
**STATISTICAL POLICY
WORKING PAPER 24**

**Electronic Dissemination
of Statistical Data**

**Prepared by
Subcommittee on Electronic Dissemination of Statistical Data
Federal Committee on Statistical Methodology**

**Statistical Policy Office
Office of Information and Regulatory Affairs
Office of Management and Budget**

NOVEMBER 1995

**Members of the
Federal Committee on Statistical Methodology**

(NOVEMBER 1995)

Maria Gonzalez, Chair
Office of Management and Budget

Robert P. Parker
Bureau of Economic Analysis

M. Denise McCormick Myers, Secretary
National Agricultural Statistics Service

Charles P. Pautler, Jr.
Bureau of the Census

Susan Ahmed
National Center for Education Statistics

David A. Pierce
Federal Reserve Board

Yvonne M. Bishop
Energy Information Administration

Thomas J. Plewes
Bureau of Labor Statistics

Cynthia Z. F. Clark
National Agricultural Statistics Service

Wesley L. Schaible
Bureau of Labor Statistics

Steven Cohen
Agency for Health Care Policy and Research

Rolf Schmitt
Bureau of Transportation Statistics

Lawrence H. Cox
Environmental Protection Agency

Monroe Sirken
National Center for Health Statistics

Zahava D. Doering
Smithsonian Institution

Alan R. Tupek
National Science Foundation

Daniel Kasprzyk
National Center for Education Statistics

Denton Vaughan
Social Security Administration

Nancy Kirkendall
Energy Information Administration

Robert Warren
Immigration and Naturalization Service

Daniel Melnick
Substance Abuse and Mental Health
Services Administration

G. David Williamson
Centers for Disease Control and Prevention

Preface

The Federal Committee on Statistical Methodology was organized by the Office of Management and Budget (OMB) in 1975 to investigate the issues of data quality affecting Federal statistics. Members of the Committee, selected by OMB on the basis of their individual expertise and interest in statistical methods, serve in a personal capacity rather than as agency representatives. The Committee conducts its work through subcommittees that are organized to study particular issues and prepare working papers presenting their findings. The subcommittees are open by invitation to Federal employees who wish to participate. This is the 24th Statistical Policy Working Paper published under the auspices of the Committee since its founding.

The Subcommittee on Electronic Dissemination of Statistical Data was formed in January 1994 to document the use in Federal statistical agencies of electronic means of disseminating data. The report, Statistical Policy Working Paper 24, *Electronic Dissemination of Statistical Data*, includes several topics, such as Options and Best Uses for Different Media, Operation of Electronic Dissemination Service, Customer Service Programs, Cost and Financing of Electronic Dissemination Service, Latest Interagency Initiatives, and an Appendix, giving a summary of current Federal agency practices.

The report is intended as an aid to statistical agency managers in decision-making about electronic dissemination of publicly-available statistical data. The report was organized and written to transcend the types of media used in the rapidly changing environment for electronic dissemination. Readers may find it more useful to begin with Chapter 5, on Cost and Financing of Electronic Dissemination Service, followed by the chapters on Customer Service Programs and Operation of Electronic Dissemination Service. The material in the Appendix and Chapter 2, Options and Best Uses for Different Media, is the most time-sensitive and will need to be updated periodically to maintain its utility. The last chapter, on Latest Interagency Initiatives, explains some "next steps" already underway for dissemination of Federal statistical data.

The Subcommittee on Electronic Dissemination of Statistical Data was chaired by Alan Tupek of the National Science Foundation.

This report and other Statistical Policy Working Papers are available electronically through the Bureau of Transportation Statistics' World Wide Web home page (<http://www.bts.gov>). A version of this report in pdf format is also available through National Science Foundation's Division of Science Resources Studies (<http://www.nsf.gov/sbe/srs/stats.htm>).

**Members of the Subcommittee on
Electronic Dissemination of Statistical Data**

Alan Tupek, Chair
National Science Foundation

Wendy Alvey
Internal Revenue Service
Department of Treasury

Judy Ball
Agency for Health Care Policy and Research
Department of Health and Human Services

Duffy Barr
National Agricultural Statistics Service
Department of Agriculture

Larry Bobbitt
National Center for Education Statistics
Department of Education

Joseph Dorsey
Social Security Administration
Department of Health and Human Services

Charlie Gilbert
Federal Reserve Board

Jimmie Givens
National Center for Health Statistics
Department of Health and Human Services

Charles Heath
Energy Information Administration
Department of Energy

Francis Horvath
Bureau of Labor Statistics
Department of Labor

John Kavaliunas
Bureau of the Census
Department of Commerce

M. Denice McCormick Myers
National Agricultural Statistics Service
Department of Agriculture

Kenneth Rogers
STAT-USA
Department of Commerce

David Saia
National Science Foundation

Jasmeet Seehra
Office of Management and Budget

John Weiner
Energy Information Administration
Department of Energy

Robert Weinzimer
National Center for Health Statistics
Department of Health and Human Services

Forrest Williams
STAT-USA
Department of Commerce

Acknowledgments

In November 1993, the Federal Committee on Statistical Methodology (FCSM), recognizing the importance of the methods used by Federal agencies to disseminate statistical information, authorized the formation of a subcommittee to prepare a report on the topic. Katherine K. Wallman, Chief of the Office of Statistical Policy at the Office of Management and Budget, enthusiastically supported the Committee's decision and hoped that the work of the new Subcommittee would encourage all Federal statistical agencies to better meet customer needs by improving the electronic dissemination of their statistical information. The Subcommittee is grateful to Katherine K. Wallman for her continuing support and her quest for one-stop shopping for customers of Federal statistics.

Each chapter of the report had a team of authors, one of whom was the chapter coordinator. Primary authors for Chapter 1, Introduction, were Judy Ball (coordinator), Joseph Dorsey, and Alan Tupek. Primary authors for Chapter 2, Options and Best Uses for Different Media, were Duffy Barr, Charlie Gilbert, Jimmie Givens (coordinator), and David Saia. Primary authors for Chapter 3, Operation of Electronic Dissemination Service, were Larry Bobbitt, Robert Weinzimer, and Forrest Williams (coordinator). Primary authors for Chapter 4, Customer Service Programs, were Charles Heath, John Kavaliunas (coordinator), and John Weiner. Primary authors for Chapter 5, Cost and Financing of Electronic Dissemination, were Francis Horvath (coordinator), Kenneth Rogers, Jasmeet Seehra, and Forrest Williams. The concluding chapter on Latest Interagency Initiatives was written by Alan Tupek. The Appendix of the report was prepared by Denice Myers. Tanya Gore, National Science Foundation, prepared and handled most of the correspondence generated by the work of the subcommittee. Wendy Alvey was the editor of the report. Bettye Jamerson, Internal Revenue Service, designed and prepared the final layout for the report.

We are grateful for the many helpful suggestions for improving the report provided by the members of the Federal Committee on Statistical Methodology. Primary reviewers from the FCSM were Charles Pautler and Rolf Schmitt. Daniel Kasprzyk also provided a great many helpful suggestions.

We feel fortunate that a number of data dissemination experts were willing to devote their time to providing valuable comments on this report. Our expert reviewers included Marty Appel, Bureau of the Census; Cathryn Dippo and Michael Levi, Bureau of Labor Statistics; Terrence Ireland; Gerri Michael-Dyer, Agency for Health Care Policy and Research; John Kirkbride, Sparks Commodities, Inc.; Mary Martin, Claremont College; Oya Rieger, Cornell University; Richard Rockwell, Inter-university Consortium for Political and

Social Research; David Roy, Statistics Canada; John T. Sabo, Social Security Administration; Joseph J. Salvo, New York City Department of Planning; Carolyn Shettle, National Science Foundation; Ed Spar, Council of Professional Associations on Federal Statistics; Timothy Sprehe, Sprehe Information Management Associates, Inc.; and Peter Weiss, Office of Management and Budget.

We also appreciate the cooperation of the statistical agency heads who provided a current (as of October 1995) assessment of electronic dissemination in their agencies. Their responses formed the basis for the Appendix of this report.

Our special thanks to Denice Myers, who made a wide-range of contributions to the work of the Subcommittee, including preparing detailed minutes of each meeting and reminding us of tasks that we had agreed to do. Thanks also to Wendy Alvey for shaping the pieces of the report into a cohesive document. Lastly, we are indebted to Maria Gonzalez, who provided overall guidance and support to the Subcommittee and encouraged completion of the report in a timely fashion.

TABLE OF CONTENTS

	PAGE
Preface	i
Members of the Subcommittee on Electronic Dissemination of Statistical Data	ii
Acknowledgments	iii
1: Introduction	1
Deciding How to Disseminate Electronically	1
Demand	1
Capacity	2
Money	2
Mandate	2
Benefits	3
Security, Confidentiality, Privacy	3
Organization of this Report	4
2: Options and Best Uses for Different Media	5
Background and Philosophy	5
Selection Criteria	5
Timeliness	6
Manipulability	6
Hardware Requirements	6
Cost	6
Size of the Data File	7
Media Choices	7
On-line Transmissions	7
Electronic Bulletin Boards	7
Facsimile	9
E-mail	10
Direct Computer Dial-in	11
Internet	12
Recorded Media	13
Optical Discs (CD-ROM)	14
Diskettes	15
Data Tapes	16

	PAGE
Data Preparation	17
ASCII Files	19
Binary Files	20
Documenting the System	21
Getting Started	22
Navigating the System	22
Technical Support	23
Metadata	23
Customer Service Support	25
Privacy, Confidentiality, and Data Security	26
3: Operation of Electronic Dissemination Service	27
Preparation for Product Release	27
Types of Data and Information Releases	27
Scheduled Releases	28
Unscheduled Releases	28
Embargoed Releases	28
Documenting the Delivery System	29
Announcement of Releases	29
Order Processing	29
Order Fulfillment	30
On-line Transmissions	30
Recorded Media	32
Scheduled Products	32
Single Orders	32
Rush Orders	33
Customer Support	33
Order Fulfillment and Accounting Support	34
Updates, Corrections, and Errata	34
Archiving	34
Back-ups	34
NARA Activities	35
4: Customer Service Programs	37
Introduction	37
Setting Up a Customer Service Program	37
Basic Activities of an Information Center	37
Assistance in Using Agency Products and Services	38
Technical Support Service	39
Agency-based Locator/Finder Services	39
The Government Information Locator System	39
Referral Services	40

	PAGE
Marketing Programs.....	40
Customer Identification and Tracking	41
Identifying Customers	41
Customer Information Systems	42
Market Research and Customer Surveys	43
Literature Searches	44
Focus Groups.....	44
Customer Surveys	44
Setting Guidelines for Customer Surveys	45
Customer Survey Instrument	46
Promotional Activities	47
Advertising.....	47
Paid Advertising	47
Unpaid Advertising	47
Exhibits and Conferences	48
Personal Contacts	48
Joint Ventures	48
Sellers, Resellers, and Redistributors	49
Federal Resellers and Redistributors	49
Commercial Resellers and Redistributors	49
Contracting-out Data Dissemination	50
State and Local Redistributors.....	50
Sources for Bibliographies and Abstracts	51
5: Cost and Financing of Electronic Dissemination	53
General Cost Estimation	53
Fixed Costs	54
Software Development and Licensing Fees	54
Hardware Costs	55
Personnel Costs for Data Development	55
Documentation of the Electronic Data Dissemination	55
Amortization of Fixed Costs	55
Variable Costs	56
Personnel	56
Customer Service	56
Billing Activities	57
Free Products and Services	57
A Caveat	57
Marketing and Promotion	58
Congestion Pricing.....	59
Guidelines	59

	PAGE
6: Latest Interagency Initiatives	61
References	63
Appendix	65
Figure A1.--Internet Addresses for Federal Statistical Agencies	66
Figure A2.--Summary of Media Availability by Agency	67
Agency Contacts and Summary of Current Practices:	
Agency for Health Care Policy and Research	68
Bureau of the Census	69
Bureau of Economic Analysis	70
Bureau of Justice Statistics	71
Bureau of Labor Statistics	72
Bureau of Transportation Statistics	73
Centers for Disease Control and Prevention	74
Economic Research Service, Department of Agriculture	75
Energy Information Administration	76
Federal Reserve Board	77
Immigration and Naturalization Service	78
National Agricultural Statistics Service	79
National Center for Education Statistics	80
National Center for Health Statistics	81
Office of Research and Statistics, Social Security Administration	82
Science Resources Studies, National Science Foundation	83
Statistics of Income, Internal Revenue Service	84
STAT-USA, Commerce	85
U. S. Geological Survey	86
U. S. Postal Service	87

1: Introduction

The rapid expansion of computer technology has led to vast changes in the supply of and demand for Federal statistical data. Technology is no longer the primary barrier between users and information. As a result, the audience for statistical data, traditionally limited to a specialized constituency, is growing and changing. The advent of a national information infrastructure will surely have even greater impact. To serve the increased demand for information, statistical agencies must adopt new methods of disseminating statistical information and data to replace the traditional means -- reams of paper and magnetic tape -- that used to serve as the principal source of statistical information.

This report is intended to aid managers in decision-making about electronic dissemination of publicly-available statistical data -- using electronic bulletin boards, faxes, e-mail, Internet and other public networks, diskettes, magnetic tape, and CD-ROM. These decisions include how to pursue new avenues of electronic dissemination of statistical data, how to operate an electronic dissemination service, how to prepare data for electronic dissemination, how to service the users, and what the costs and financing options are for electronic dissemination. The report includes recommended procedures for electronic dissemination of data -- e.g., suggestions on user charges -- and ideas for dissemination of survey concepts and measures of data quality -- as well as examples of current practice and near-future plans of Federal statistical agencies in the area of electronic dissemination of statistical data.

It should be noted from the outset, however, that this field is changing daily. The focus here, therefore, is on **process**, rather than on recommendations of particular technological systems. Examples, when given, provide a baseline for current best practice.

Deciding How to Disseminate Electronically

A variety of factors will contribute to an agency's ability to begin and sustain an electronic dissemination function. This section highlights six areas for consideration by agencies in deciding how to initiate electronic dissemination.

Demand

Users of government data are asking for more data, more quickly, in a more user-friendly form. Electronic dissemination affords a method -- in some instances, the only method -- of meeting these increased demands. In fact, some agencies have discovered that moving to electronic dissemination (e.g., from paper reports) or moving into a more user-friendly form of electronic dissemination (e.g., from magnetic tape to CD-ROM) not only serves current users of their data, but yields more users, as the data become more easily accessible. Ultimately, electronic dissemination should prove to be less expensive for organizations which today have extensive print publishing activities, given the high cost of paper, printing, distribution, order taking, and warehousing.

Capacity

Adopting an electronic dissemination strategy requires an agency to be willing and able to invest the resources to create and disseminate data in this new way. Capacity, of course, may encompass many things, including:

- **Personnel.**-- Existing personnel may need retraining and, possibly, new personnel with new sets of skills may be necessary;
- **Equipment.**--An investment in hardware and software may be required to permit the agency to generate data or documents in electronic form;
- **Contracts.**--Rather than investing in personnel and equipment in-house, the agency may choose to contract for the creation and dissemination of electronic data;
- **Dissemination Service.**--The agency will need to decide how to distribute its electronic products to users, again through in-house personnel or by contract;
- **Customer Support.**--As a corollary to distribution, the agency will need to provide customer service and technical assistance, if the agency does not contract out its electronic dissemination. Whereas traditional assistance tends to focus on users' questions about content, the customer support for electronic products may also need to handle the myriad of users' technical difficulties that result from incompatible hardware/software and/or users' ignorance about technical aspects that arise during the use of an electronic product -- e.g., why does the software that accompanies these data on CD-ROM crash my PC?

Money

Regardless of the agency's capacity to staff various aspects of an electronic dissemination function, the agency may also need to cover costs other than personnel, either by reallocating current budget dollars or by identifying new sources of funding (e.g., user fees). Garnering realistic estimates of start-up and continuation costs is clearly important.

Mandate

A requirement by the Executive Office of the President, the Office of Management and Budget, or the Congress may represent compelling reasons for an agency to decide to disseminate data electronically. In addition, the National Performance Review urges agencies to increase efforts to deliver information and services to the public via electronic communication. Agencies can turn such mandates into opportunities by anticipating the direction technology is heading and planning strategic moves into electronic dissemination.

A less well-known mandate may come from the requirements of the National Archives and Records Administration (NARA). Agencies are required to provide data or information meeting the NARA criteria of "permanent records" for entry into the permanent archives of

the United States. Whether NARA requirements affect the choice of electronic dissemination strategy will depend to a large extent on the schedule for retirement of data and/or information. Today's technologies are likely to be applicable only for records scheduled for retirement in the near future.

The Government Information Locator System (GILS) is a government-wide program to help the public locate and access public information. GILS will also help standardize the process of identifying and describing public information resources.

Benefits

Electronic dissemination offers new opportunities for statistical agencies. For example, well planned and executed electronic dissemination strategies may release resources currently used to handle data requests. Electronic dissemination may also result in more efficient means of providing updates, including, in some cases, much earlier releases of data that are frequently revised.

Electronic dissemination also provides opportunities for agencies to rethink what is the best way to convey information. Traditionally, the presentation of information about the reliability of data has posed special challenges to statistical agencies. In printed reports, space and format issues limit the presentation. Electronic reports offer more options. For example, electronic connections can be created to link statistical reports and their underlying methodologies, which have traditionally been available only as separate reports. Statistical reports that reference other sources of data can be linked electronically to the methodologies for the source data. In addition, electronic methods can provide easier access to standard error estimates or permit the calculation of such estimates.

Security, Confidentiality, Privacy

Agencies that deal with sensitive data, especially data that may be used to identify individuals (e.g., health care events), must consider how security will be preserved and confidentiality and privacy protected in an electronic dissemination environment. The ease with which electronic files can be combined adds to the potential threats to security. Data that are non-sensitive when in paper format or in independent files may become quite sensitive and reveal identifying information when combined electronically with other available files. In addition, electronic access inevitably raises the specter of **unauthorized** access. Barrier methods are needed to ensure that customers using publicly-available data cannot find their way into nonpublic files. Because of the special problems posed by sensitive data, this report will focus primarily on public-use information. Still, although data protection issues are largely beyond the scope of this report, agencies must expect electronic dissemination to pose new challenges in these areas.

Organization of this Report

This report is divided into six chapters and an Appendix, which draw on the experience of state-of-the-art systems currently in operation or being planned by Federal statistical agencies. The remainder of the report covers the following:

- **Chapter 2: Options and Best Uses For Different Media.**-- As the world becomes more interconnected electronically, agencies are likely to select electronic means as their first choice for data dissemination. This chapter describes different media to consider, with a focus on timeliness, manipulability, hardware requirements, cost, and size of data files.
 - **Chapter 3: Operation of Electronic Dissemination Service.**--This chapter covers the preparation of certain types of statistical information for electronic release, order processing and fulfillment, customer support and archiving issues. The importance of advance planning is emphasized, and data preparation tasks are discussed.
 - **Chapter 4: Customer Service Programs.**--Customers and customer service are at the heart of the Federal government's programs to implement quality management and achieve high standards of customer satisfaction. This chapter stresses the need for well-trained personnel to respond to customer requests for information and describes the components of a successful operation, from product documentation to software and adequate hardware facilities.
 - **Chapter 5: Cost and Financing of Electronic Dissemination.**--Traditionally, government information providers have relied on printed documents with a set of associated costs. Both fixed and variable costs for providers and users of electronic information are different. This chapter discusses the costs associated with the electronic dissemination function, as well as important issues to consider in accounting for expenses and equity considerations necessary to recover costs from the user.
 - **Chapter 6: Latest Interagency Initiatives.**--There already are a number of new efforts underway to expand electronic dissemination of Federal statistical data. This section lays out the "next steps" for the near future.
 - **Appendix:** This section provides an agency-by-agency summary of current practices in electronic dissemination of statistical data. Although not intended as a comprehensive survey of Federal dissemination efforts, the Appendix covers current activities of many of the major statistical agencies, including contact information for selected data products.
-

2: Options and Best Uses for Different Media

Background and Philosophy

Publicly-available statistical data have been produced and disseminated for years in paper form (hardcopy) -- even magnetic tape has often been accompanied by printed codebooks and reams of printout. Producers of hardcopy were encouraged to abide by rules of the press for typesetting, table formatting, and spacing. Many of these requirements were implemented simply to improve the document's visual aesthetics and readability.

Today, we are faced with the challenge of preparing and distributing data and documentation in electronic form. Some of the knowledge garnered through many years of paper distribution will no longer apply. Instead, there are new specifications to ensure that the electronic information remains readable, accessible, and acceptable by customers. Of course, few would disagree that there are some decided benefits that result from the move to more modern electronic media. Prominent efficiencies include:

- increased access to data by a broader variety of users;
- faster transfer of data from producer to users;
- more efficient use of human resources involved in answering data requests;
- reduced burden on users to re-enter data, especially when large datasets are involved; and
- expanded ability of the user to analyze data.

While striving to maximize the benefits received from electronically distributing data, there must also be a balance between the potentially diverse needs of customers and the agency's own resource limitations. As the producer of statistical data for dissemination, agencies must know who the customer is and, thus, target data products for that audience. There are a variety of tactics to determine who uses the data products and what uses are made of them. These tactics range from conducting market research surveys to compiling lists based upon historical data-use information already available to the agency. Most data products must meet the needs of a variety of customers. Meeting these needs will require robust products which are easily available, usable, and affordable.

The rest of this chapter will describe the most prominent and promising media choices available today, in terms of key selection criteria. It will then go on to discuss some general issues regarding data preparation and system documentation, which go hand-in-hand with those choices.

Selection Criteria

Before selecting a medium for disseminating public-use data electronically, questions must be answered which deal with timeliness, manipulability, hardware requirements, cost,

and size of data file. Once these criteria have been considered, the data producer can better decide which medium best suits the needs of both the data environment and the user.

Timeliness

Agencies must first determine how soon the data are required by primary customers -- within minutes, hours, or days after release -- and frequency of data retrieval. Another timeliness issue is how quickly electronic data can be made available to the public after data collection. Different methods of data transmission can be considered, such as phone, facsimile (fax), and electronic transfer over modems. Less timely methods for less time-sensitive transmissions may involve the use of mail or postal express, where CD-ROMs and floppy diskettes can be shipped to users. If greater timeliness is important, agencies will likely want to choose data collection and transfer methods that do not require traditional physical handling and transportation.

Manipulability

Ease of use is another major consideration. Customers analyzing the data may prefer a format that permits direct manipulation without need to re-enter the base data -- e.g., data that can be easily imported into a spreadsheet or some other application. Manipulability also encompasses the ability to create smaller files using only a subset of the information. Subsetting can be accomplished through user selection of specified variables; for example, some analysts may only want to look at a particular state's or region's data from a file that contains all states and regions. As the goal of providing greater access to a broader group of users is achieved, it will become more complicated to meet their data manipulation requirements -- because of differences in their interests and expectations, variations in their technological expertise, and increasing concerns about potential problems with maintaining privacy and confidentiality.

Hardware Requirements

Another issue to bear in mind is hardware requirements. Since it is desirable to make data available to as wide an audience as is practical, it is a good idea to provide data files that are accessible by the majority of the users' hardware. In order to do so, it may be necessary to query customers about their hardware systems and configurations. (If you choose to produce data which are hardware-restrictive, there should be an excellent reason for it or a more universal alternative offered for similar data.) When contemplating the hardware, also consider the operating systems -- for example, OS/2, Unix, and MacIntosh. It is also important to remember that hardware continually changes and it may be difficult for users to remain current. While agencies implementing electronic dissemination programs surely strive for state-of-the-art solutions, these systems may be incompatible with the user's existing hardware. Hence, provisions should be made to convert the data files to readable form.

Cost

The cost of electronic dissemination is a major decision-making issue, as well. There are two cost issues that deserve early consideration:

- the initial cost for start-up and
- the maintenance cost associated with keeping the system running.

Bear in mind that, since one goal of Federal agencies is to provide as much data as possible at the lowest possible cost, the cost to both the agency and the user must be factored into deliberations when preparing data for dissemination. For instance, it is crucial that the cost for data access not be so high as to cut off a major portion of the known and potential customer base.

Size of the Data File

Finally, size of the data file is an important factor in choosing the dissemination media. Large datasets generally take longer to plan and produce. They may also employ physical media forms, such as a CD-ROM or a magnetic tape, that require physical transportation to the users, lengthening the data transfer time. Smaller files, on the other hand, provide greater flexibility with respect to choice of media and availability, since they may be either electronically transferred -- in on-line fashion, using file transfer tools -- or physically transported via a variety of media types.

Media Choices

This section describes each of the more commonly-used media forms, focusing on each of the criteria specified above. These have been grouped into two key types of data storage methods currently being used in Federal statistical agencies -- on-line transmissions and recorded media. Clearly more than one type of media may be appropriate when establishing a data dissemination program.

On-line Transmissions

As the world becomes more interconnected electronically, electronic transmission of data is likely to become the primary means of data dissemination. This category of media is especially strong in fulfilling users' needs for timeliness and affordability. Below is a brief summary of the major methods of "on-line" or electronic transmission in use today.

Electronic Bulletin Boards.--This forum for electronic dissemination was the pioneer form for communicating data "on line." With growth of mainstream on-line services like *CompuServe* and *America On-Line*, users were able to depend less on electronic bulletin boards as their sole source of easily accessible data. In view of the recent popularization and exponential growth of the Internet, this trend is likely to continue. Still, the electronic bulletin board may be a good way to start disseminating data on-line.

Bulletin Board Services (BBS) -- once thought of as viable alternatives to most on-line systems -- are a low-cost means to provide users with access to agency public-use data. A few advantages of a BBS are that:

- setting them up is not expensive;
 - there is generally no cost to users; and
 - they require users to have only the most basic hardware setups.
-

Specifically, a large number of users can access them because the access is made available via lowest-common-denominator computers, modems, and telephone lines. The approximate transmission rate for file transfers using an Xmodem protocol and a 9600 baud modem is 7.5 minutes per megabyte (MB). The BBS is often used for text, tabular materials, and, occasionally for small microdata files. The BBS is also a medium for many discussion forums, where users engage in electronic dialogues by leaving and receiving messages about topics of specific interest. The following details may help in considering installation of a Bulletin Board Service.

- **Timeliness.**--Data can be uploaded by a systems operator at any time, so that the most up-to-date data may be accessed by users. The only limitation is the frequency with which the operator is able to upload files.
- **Manipulability.**--Files are generally posted in a form that can be easily manipulated by users.
- **Hardware Requirements.**--In order to set up a BBS at least one dedicated telephone line, a modem or modem pool, and a computer are needed. The user minimally must have access to a modem, a telephone line, and a computer of any type. (Note that by posting public-use data to a stand-alone computer, the data producer can protect the source data from illegal intruders.)
- **Cost.**--Set-up costs for a BBS are probably the lowest of those for all avenues of electronic data delivery systems. Usage costs are also low, especially for users in the local calling area. Most BBS require that long distance users incur the cost of a long distance telephone call; there are a few exceptions, though, where the agency provides an 800 number. Apart from telephone company charges and the customer's own computer equipment, however, BBS use is essentially free. Once the infrastructure is in place, the data producer's costs are limited to those associated with system maintenance, i.e., to keep the system current.
- **Size of Data File.**--This consideration poses the greatest limitation to selection of BBS as a data dissemination medium. Because users are downloading files, and all but local users are charged for line use by the minute, this forum is most suitable to smaller data files -- i.e., files of less than one megabyte. Many BBS also have time limits for individual calls, such as 45 minutes per call or 60 minutes per day. Because of the premium on time used in downloading, it is customary to compress data files posted on BBS. The most widely used compression format is *pkzip*. Most BBS systems post a version of *pkzip* for users to download and use.

ADVANTAGE: This mode of dissemination assures a high level of user access, since the hardware and software requirements are minor and likely are already available to most data users.

DISADVANTAGE: Since this dissemination mode is so readily accessible, it becomes imperative that adequate access lines be available and that the software servicing the BBS be easy to use.

Facsimile.--It appears that no office is complete without at least one fax machine. Faxes have become an integral part of everyday operations. Although fax technology has the advantage of being able to transmit an image of any printed page, typical fax machines are considered too labor-intensive for use as a primary means of distribution. The Automated Fax-Back (AFB) technology, however, may be a viable alternative for some agencies. Modem manufacturers have greatly improved technology by incorporating fax capabilities into modems. Conversely, some fax manufacturers now incorporate modems into their faxes. Using an AFB system, a caller follows an automated voice menu, which typically provides details about available data. The system also provides instruction on selecting the desired information and arranging that the data be returned by fax to a specific individual at a caller-designated fax telephone number. An automated system then sends the requested data to the fax machine that has been designated by the caller. Fax technology also provides the ability to automate the routine distribution of faxes. This feature is important for agencies that periodically distribute data to a lists of users. Agencies may consider the facsimile method of electronic transfer an excellent supplement to their primary information distribution mechanism. When weighing the facsimile alternatives, consider the following before deciding that this is the best solution.

- **Timeliness.**--As with other media, data can be as current as the systems operator can post them.
 - **Manipulability.**--Fax offers no potential for manipulability. The final data product is merely an image that exists on paper only -- it is read-only.
 - **Hardware Requirements.**--Both the data producer and the data user will need a fax machine and telephone, which are standard office equipment. The more sophisticated systems will be software driven and require a computer with reasonable storage capacity to contain the public-use information which the customers can request.
 - **Cost.**--Set-up costs will vary. The data producer's agency cost will vary, also, depending upon the type of system implemented. Naturally, the fully automated system with both an incoming and outgoing call modem pool will be more expensive than the single line non-automated system. The additional expense, however, may well be worth it, if a primary concern is providing adequate customer access to information. Consumer costs will vary from as little as the cost of the phone call to as much as the cost of a computer with a fax modem and fax software. If hardcopy capability is also desired, the user will need a printer, as well -- also standard office equipment.
 - **Size of Data File.**--In general, one would not fax more than a dozen or so pages. Therefore, this medium is most suited to single-page tables and the like. In most cases, limitations on number of pages faxed is not because of internal hardware limitations, but, rather, result from the fact that customers often have difficulty receiving large documents. Another restricting factor is that fax systems will re-size graphic images, further reducing the number of pages which can be successfully transmitted.
-

ADVANTAGE: This mode of dissemination assures a high level of user access, since fax machines are commonplace in most office settings. Individual computers with a fax modem and software are also capable of accessing these systems.

DISADVANTAGE: This method is closely akin to current paper distribution systems and can be quite cumbersome.

E-mail.--Analogous to the automated fax-back system are systems that receive electronic mail (e-mail) requests for data and automatically send the data requested back via e-mail. Although addressed in this report as a separate entity, e-mail is really an external dissemination option which enhances the Internet connection. Variants of this method of electronic dissemination have been in existence for years before being made popular by the advent of the Internet. Similar to fax technology, it is also possible to automate the routine distribution of data by electronic mail. Agencies that periodically distribute data to a lists of users with e-mail capability may find this feature useful. In reviewing e-mail for data dissemination purposes, consider the following.

- ***Timeliness.***--Given that the implementation of certain e-mail systems require that messages travel rather circuitous routes between sender and recipient, this kind of system is unlikely to be able to satisfy users who need extremely time-sensitive data immediately upon release. However, most systems will provide data on the same day that it is requested -- as opposed to within minutes of the request. Therefore, agencies setting up mail back systems should inform customers of this limitation; otherwise, they may find themselves involved in troubleshooting response time issues.
 - ***Manipulability.***--E-mail is one of the more manipulable forums available. The user may be able to make very specific requests if the data producer designs the system with that intent. Furthermore, the files received by the user are likely to be in a form that can be easily manipulated by the user.
 - ***Hardware Requirements.***--Both the data producer and the customer must have computers, modems, and e-mail access. This option is very cheap for the disseminator if the volume of requests is not so great that a dedicated machine is required. No phone lines are needed beyond the one normally used to receive and transmit e-mail. In instances where volumes are large, it may be necessary to install a dedicated mail server or router with sufficient storage capability.
 - ***Cost.***--Data transmissions via e-mail require minimal additional expense beyond regular e-mail costs. Additional cost will be incurred in situations where additional hardware becomes necessary or when it is determined that the existing dataline band width is inadequate for the anticipated volume of traffic.
 - ***Size of Data File.***--The size of the file disseminated may be limited by e-mail hardware considerations. Whether pre- or post-Internet, e-mail systems have largely been
-

used as a method to rapidly transfer text-based information, i.e., data and reports. These kinds of files do not generally exceed one or two megabytes and are frequently much smaller, so their size poses no real concerns.

ADVANTAGE: This method offers on-line access to users -- like some of those in the Federal government -- who may not yet be able to use the Internet.

DISADVANTAGE: Some users have difficulty reading e-mail documents that were created with software which is different from their own. This is especially true with larger datasets and documents.

Direct Computer Dial-in.-- This data dissemination method permits users to log on to a computer and directly access software to manipulate publicly-available data on the remote machine, including subsetting datasets, performing calculations for data analysis, and formatting data in the best way for the user. This alternative is similar to a Bulletin Board Service, except it gives the user more flexibility to select specific items of interest from a larger dataset; bulletin boards and their equivalents usually have data in files, put together in what the distributor perceives will likely be the most popular groupings. Access to the dial-in database may be controlled from a series of menus to make it easy for users to select the information they require. Depending upon the available software tools, this forum will allow immediate access to text and summary data in graphics or tabular format. Within the Internet, the Telnet command gives a user this form of access. In evaluating whether a direct computer dial-in system best meets your requirements, consider the following.

- **Timeliness.**--Many of the factors that affect the timeliness of an electronic bulletin board have similar effects on the potential timeliness of computerized direct data access. Updated files can easily be made available quickly, though perhaps not fast enough for some very time-sensitive data. Furthermore, direct dial-in has the potential to improve speed significantly if only some of the data on a file are needed, e.g., by subsetting, the user can specify only those items desired and, thus, speed up the overall retrieval process.
 - **Manipulability.**--Users who access data through direct dial-in get two chances to manipulate their data, once on the machine they are accessing and then, again, back on their home machine after retrieving it.
 - **Hardware Requirements.**--The direct dial-in hardware required by both users and producers is similar to that required for a BBS. In addition, with users manipulating files, the data disseminator will probably need more processing capacity and disk space.
 - **Cost.**--Both user and producer costs are comparable to those for the BBS. While the cost of developing routines to extract the data desired may add to the expense of the system, it is likely to be offset by the decrease in time that would have been required for the user to transfer a larger file back from a bulletin board.
-

→ ***Size of Data File.***--The data file extracted has similar size constraints as a bulletin board file.

ADVANTAGE: This mechanism provides greater flexibility in manipulation and transfer of data than the BBS.

DISADVANTAGE: System security and the ability to maintain confidentiality of data will be challenged. The data access interface must be intuitive; otherwise, technical support requirements will increase.

Internet.-- While, assuredly, the most talked about and written about method of electronic information delivery, the Internet has actually been around for quite some time. It is a system of computer networks which was begun by the Department of Defense to ensure a national and allied communications capability. The Internet was later refined by the National Science Foundation. Since then, it has expanded exponentially, attracting users far outside of the original core of researchers and military and government personnel. Today, the Internet has evolved into the world's largest network, one often referred to as the "Information Superhighway." Currently, more than 25 million users of the Internet may tap into information available in more than 45,000 computer networks (e.g., research labs, government agencies, universities, libraries, museums and many public school systems) in nearly 200 countries. Accessing the many available islands of information, users can transport text, graphics, and even audio and video files directly to their own computers. Furthermore, mainstream on-line services continue to incorporate more aspects of Internet connectivity into their basic packages. Although the Internet has proven an invaluable resource, some novice users find successfully navigating the Internet can be a tedious undertaking. Because the Internet is growing and evolving so fast, this discussion will be limited to those basic aspects of this global system of networks that are not as likely to change. When considering data dissemination on the Internet, the following information may help you make an informed decision.

→ ***Timeliness.***--As with BBS, the only limitation to timeliness is the frequency with which a contributor is able to post files. This resource is valuable not only because it provides near immediate access to recent data, but also because it allows access and retrieval of historical data.

→ ***Manipulability.***--Just as with the BBS, files are generally posted in a form that can be easily manipulated by users. Because of its complexity and size, a user able to successfully browse the Internet may find very recent postings of data on one or several of its nodes. The user need not know where to look, however -- simply how to look. This method of electronic dissemination provides a great measure of versatility, particularly when linked to the World Wide Web -- the fastest growing part of the Internet. With "browsers" or client software -- e.g., *Netscape*, *Navigator* or *Mosaic* -- users can preview graphic and image data while on line. Other Internet tools available include Telnet, Trumpet, and File Transfer Protocols (FTP) to access on-line data. Files may be printed or stored at the customer's site for later use with an analytical tool of choice.

- **Hardware Requirements.**--Direct access to the Internet is well within reach of most ordinary computer users. Both the data producer and the data user will need only a microcomputer, a modem, and a Serial-Line Internet Protocol (SLIP) or Point-to-Point Protocol (PPP) Internet. Unlike other direct connections to the Internet, which are usually by means of a high speed data line, such as a T1 line, the SLIP/PPP connect may be achieved using normal telephone lines. Most direct Internet connections are to Unix workstations that are directly networked to the Internet. As mentioned above, however, approaches that connect to the Internet via commercial services -- such as *Prodigy*, *America Online*, *CompuServe*, or specialized Internet access providers -- are now growing in popularity. The customer uses a modem via telephone line to access a commercial workstation; the workstation or "server," in turn, connects directly to the Internet.
- **Cost.**--Customers are acquiring SLIP/PPP accounts at monthly rates similar to those charged for basic monthly telephone service. Customers using modems may connect for a few dollars less per month, but generally pay an additional hourly access fee. Note that the Internet is evolving with breathtaking speed! These costs, therefore, are likely to change, especially as economies-of-scale begin to take effect.
- **Size of Data File.**--Predicting size limitations for the Internet is difficult, particularly because of the rate of technological change as Internet use becomes prevalent. Assuming a data transmission over a T1 line at 1.544 megabits per second, transmission of large data files is not unthinkable. File size limitations for this mode of dissemination are constrained only by pragmatic factors. For the producer, these factors include -- but are not restricted to -- internal data storage capacity and available band width for file transfers. Customers may also find themselves limited by line charges associated with long transfer times, hardware ability to read large files, and data storage capabilities.

ADVANTAGE: The Internet access mode gives a growing number of users an unprecedented ability to obtain data and to manipulate large datasets.

DISADVANTAGE: A greater need for technical support may arise because uninformed users will access the data and will require explanations as to what the data represent. Similarly, users with limited knowledge of the agency, will require assistance in using the systems they access.

Recorded Media

As a precursor to electronic on-line transmissions, data on disk or tape have been available for some time. These media are still useful, as they allow for the transmission of large volumes of data. They have, however, longer preparation and delivery times than other media discussed in this report.

Optical Discs (CD-ROM).-- Compact Disc/Read-Only Memory technology -- or CD-ROM, as it is commonly known -- is most suitable for dissemination tasks requiring that large amounts of data be made available to users for whom timeliness is not a major consideration. As opposed to files downloaded from the Internet or from a BBS, that are generally limited to only a few megabytes, a single CD-ROM disc can easily hold 600 MB of data. The vast storage capacity of these discs also permits producers to include data in several formats and to provide customized search engines, along with information on how to use the data -- including metadata (or data about the data).

This medium is often referred to as a write-once-and-read-many-times device -- originally an accurate description of the media, because once the disc was "written," it could only be read. Recent technology is changing this, making these discs rewriteable. At this time, it is not clear how this feature will influence decisions about the media type. It appears, however, that the ability to write a CD-ROM does provide the producer an inexpensive means of creating master discs that can later be mass-produced. As agencies deliberate about electronic dissemination systems, the following information on optical discs should be considered .

- ***Timeliness.***-- As with traditional publishing, CD-ROM production should not be considered when users need the data quickly. Because of the longer production lead-time, it is not a suitable medium for very time-sensitive data.
 - ***Manipulability.***--Because files can be posted in several formats on CD-ROM, users can easily copy them and input them into a spreadsheet or other data manipulation software. Producers can even include data manipulation software on the disc. The recent movement towards rewriteable CD-ROMs adds a new dimension to this medium's utility. Users will have the option to add their own data to these discs, making them a more valued product from the customer's perspective.
 - ***Hardware Requirements.***--Although it is possible to purchase equipment to mass-produce CD-ROMs in-house, at this writing such an expense may not be advisable. However, prices for the hardware are falling rapidly. Cheaper hardware should reduce costs, even if you need personnel to "run" the equipment. Currently, however, it is still probably more cost-effective to have the disc produced by an external CD-ROM publisher. Therefore, the only hardware a data disseminator needs is a computer; the publisher produces the disc. Agencies contemplating internal publishing will need, in addition to mastering hardware and authoring software, appropriate equipment and materials for mass reproduction. Users need a CD-ROM drive, an accessory which is becoming more and more commonplace in home personal computers, because of the popularity of multimedia options. Users without them will find that most public libraries have CD-ROM readers.
 - ***Cost.***--Costs are expected to decline as use of this technology becomes more widespread. Agencies' costs will be determined by whether or not that discs are published internally or via contract. Consumer cost will vary from one agency to another, de-
-

pending upon the agency's cost recovery program and upon whether the agency or the customer pays the shipping fees. Several distribution options may also affect user cost -- e.g., distribution may be free of charge, discs may be sold individually, or discs may be discounted when sold as a set or a subscription to a set.

- ***Size of Data File.***--As much as 600 MB of data, including search engine and technical information, can fit on a single disc. This especially large capacity is the strongest appeal of the medium.

ADVANTAGE: The greatest advantage of CD-ROMs is their large capacity.

DISADVANTAGE: Because they are recorded media, CD-ROMs are generally considered less timely than on-line data can be.

Diskettes.-- Diskettes are used most often for transporting data between two computers that are not otherwise connected. They are also the medium of choice for packaging software. Use of diskettes, however, does offer several disadvantages when compared with electronic transmission. Unlike files flowing across cyberspace, diskettes can be lost, stolen, or damaged. The primary means of disseminating information on diskette is through physical processes. Nevertheless, because diskettes currently are more accessible to users than are data available via modem, CD-ROM, or tape, they remain a popular option -- especially for mid-sized tasks. They are suitable for text files, tabular information, microdata, and graphics. Today's typical 3 1/2-inch disk has a capacity of 1.4 megabytes; high density 5 1/4-inch disks -- no longer widely used -- have a capacity of 1.2 megabytes. By comparison, a single CD-ROM disc can hold data equivalent to about 600 high-density 5 1/4-inch floppy disks, but a singular floppy disk can hold data generally considered too voluminous for easy transmission across telephone lines. When evaluating diskettes for data dissemination, bear in mind the following details.

- ***Timeliness.***--Diskettes must be formatted (or purchased in pre-formatted form), copied, packaged, and physically delivered in order to be used for dissemination of data. These tasks generally require longer lead-time than preparing data for on-line transmission -- a reflection of mass production requirements. When mass producing a disk or CD-ROM, typically a master is first created -- the effective equivalent of the final form of the electronic dataset. It can, then, be mass produced and distributed easily. That process, plus the physical transportation required of diskettes, like CD-ROMs, makes them less suitable for very time-sensitive data releases.
 - ***Manipulability.***--Diskettes generally contain files that can be easily manipulated by users. They also require software -- it may either be supplied by the customer or by the data producer and, in some cases, is contained on the data disk.
 - ***Hardware Requirements.***--Users will need a platform that has access to a floppy disk drive. With increasingly rare exceptions, almost all computer users have at least one floppy drive -- some, however, may only be able to read one type of diskette, either 5
-

1/4 inch or 3 1/2 inch. Like CD-ROMs, agencies considering mastering disks will need appropriate hardware, should they decide to produce the disks in-house. This, however, is far simpler than the equipment required to master CD-ROMs.

- **Cost.**--The per-box costs of a box of diskettes is a minor expense, not unlike purchasing a few reams of copier paper. The major costs associated with these systems are those costs incurred in producing the master copy; establishing and maintaining an in-house duplication system, and shipping or mailing the product to users. An alternative is to contract out duplication and shipping, which may reduce production expenses, but could increase user costs.
- **Size of Data File.**--Diskettes are most appropriate for medium-sized data files. The best use of diskettes as a distribution medium would be to limit a given package or dataset to a small number of disks. The fewer disks involved in downloading the data, the better. This appears to hold true, regardless of whether the user is retrieving data, restoring a backup, or installing software. In response to this concern, software houses have resorted to using compression techniques to reduce the number of disks required for storage and application. Diskettes can be made to store more data by using data compression techniques. Depending upon the data format -- i.e., graphics versus text -- through data compression as much as 5MB of data may be stored on a single 3 1/2 inch disk. (Although it is possible to apply these compression techniques to data stored on CD-ROM, most agencies will find that this is not necessary, since current technology will permit a single CD-ROM to hold up to 600MB uncompressed.)

ADVANTAGE: Diskettes are, perhaps, the most universally acceptable media for electronic data dissemination.

DISADVANTAGE: The 5 1/4 inch disks are no longer as widely used and, as more and more computers become available with CD-ROM drives, even the smaller diskettes may become obsolete.

Data Tapes.-- Perhaps the most familiar medium for data dissemination, the data tape is most suited for storage and transportation of large datasets, such as public-use microdata files. The time and expense required to produce edited data for tape are generally substantial. However, like some other media, this medium can accommodate users with a wide variety of software. Tapes are also versatile in that they may contain data in several formats, including text, microdata, summary tabular information, SAS datasets for IBM, flat files, SAS control cards used to create a SAS dataset for non-IBM systems, and SPSS-X control cards that can be used to create an SPSS-X data file from a flat file. The following issues are important in considering data tapes for public-use data dissemination.

- **Timeliness.**--Preparing the microdata for public access usually takes weeks or months, although the actual creation of the tape is not a lengthy process. Furthermore, tapes, like other recorded media, must be physically transported; hence, they are not suitable for highly time-sensitive data.
-

- **Manipulability.** --This medium's large data storage capacity (up to 5 Gigabytes) allows for data to be provided in a variety of formats. Data stored on tapes are one of the oldest mechanisms available and have traditionally been used in mainframe environments. Many Unix and microcomputer systems are now capable of accessing peripheral devices that will read or write to tapes, although that ability may not be as thoroughly developed as on a mainframe. Like the smaller diskettes and more modern CD-ROMs, the stored data can generally be easily manipulated by customers, using software supplied by either the user or the data producer.
- **Hardware Requirements.** --To prepare files on tape, the producing agencies will need hardware such as a mainframe, a Unix system, or a personal computer with at least one magnetic tape device. Users must have access to a mainframe computer or some other computer with a tape device. As this is older technology, most agencies and many regular data users already have those capabilities, but smaller, newer customers may not. Mass reproduction of tapes is probably more costly than the cloning of CD-ROMs or diskettes.
- **Cost.** --The mechanics of producing a data tape are not a costly endeavor. The costs, like those of the diskettes and CD-ROMs, are influenced by decisions about mass reproduction. If hardware and personnel resources are already available to produce the master copy, it is not difficult or costly to produce a limited number of copies. The cost for few copies is small, in that it covers only the material and shipping. When reproducing large quantities, agencies will incur not only the material and shipping costs, but will have either expenses associated with mass reproduction hardware or contracted services costs for externally-produced products.
- **Size of Data File.** --As stated above, data tapes can store very large amounts of data. A cartridge tape can contain as much as 5 Gigabytes (GB) of data. The typical 9-track, 6250 bytes per inch (bpi) tape, at maximum capacity, will hold approximately 168MB of data. So, data tapes are clearly the most suitable choice when disseminating very large files, especially when the files are very large and transfer rates are an issue.

ADVANTAGE: This medium is best for very large datasets and accomodates a full range of formats on one tape.

DISADVANTAGE: Smaller users -- like newer customers who are learning about Federal statistical data through on-line services -- may not have the hardware capability to use data tapes.

See Figure A on the next page for a summary of the strengths and weaknesses of each of these media types.

Data Preparation

Once an agency has chosen a medium for data dissemination, it must decide how data will be presented on that medium. The basic goals in this process are to make data available to

Figure A. -- Summary of Options for Different Media

Ranking	Timeliness	Manipulability	Hardware Requirements	Cost to Set Up	Suitable for Large Data Files
Best	Bulletin Board Direct Dial-in Internet Fax	CD-ROM Data tape Direct Dial-in	Diskette	Bulletin Board Direct Dial-in	CD-ROM Data tape
Very good		Bulletin Board Diskette Internet	CD-ROM Fax	Diskette Fax	Direct Dial-in Internet
Average	Diskette		Bulletin Board Direct Dial-in	Data tape Internet	Bulletin Board
Poor	CD-ROM Data tape	Fax	Internet Data tape	CD-ROM	Diskette Fax

anyone who wants them and to make it as easy as possible for the users to manipulate the data once they are received.

One important aspect to producing data files is to know what sort of processing users have planned for the data. For example, if users might import the data into a spreadsheet, the data file should be designed with that in mind.

When designing a data product, remember that users are likely to have a range of different equipment and skill levels (i.e., their knowledge of hardware and software, its use and configuration). Because most of the media discussed above transfer data to the user in a specific format (or, if their capacity is large, several formats), the producer needs to consider whether to design the data file for the lowest equipment and skill levels in the likely audience. When producing general-use data, which have widespread application and, perhaps, a shorter shelf-life, that might be a good idea. When designing a product for a more specialized user, the dissemination system may need to incorporate more sophisticated technology. At any rate, under most circumstances, more sophisticated users will adapt to your data presentation format, so honing the data to the novice user may be a good idea. Remember to bear in mind, however, that this strategy may not be appropriate for agencies desiring to ensure that their system accommodates future technological innovations. (See Chapter 4: Market Research and Customer Surveys.) Other issues to consider are the requirements for long-term storage and archival of data.

There are two basic format types currently in use -- ASCII and binary files. With this framework in mind, we can discuss some of the specific issues to consider when choosing a format. In time, as software and hardware technologies advance, many of these consider-

ations will probably become obsolete. Certainly some current restrictions -- such as file size -- will change as technology improves. Ongoing user surveys are suggested to ensure that the data producer stays informed of users' changing needs.

ASCII Files

The most generic format available to a data producer is an ASCII file. These files can be browsed and understood using any command that displays the contents of a file onto a character-based screen. ASCII files can attempt to mimic the contents and format of a hardcopy release. They may also be used to convert files to a format for further processing. The generic nature of ASCII files makes them a very democratic means of distributing data, since nearly everyone who loads them onto a computer can view and print them. In addition, the use of ASCII also minimizes the degree of users' support necessary.

Two important liabilities of ASCII files are:

- They almost always require added effort for the user to load them into software, compared with data already formatted for the customer's particular software.
- These files are relatively large, compared with binary files that contain the same information but require file translation before being read by specific software.

Both of these limitations can be alleviated somewhat by sensibly designed files. Certain conventions can allow popular software to import ASCII files more easily. Software is readily available that compresses files using algorithms that, among other things, reduce repeated character sets. The best known compression utility is *pkzip*; it can either create an executable file that will explode itself into the uncompressed file or it can create a compressed file that is uncompressed using *pkunzip*. Executable files that uncompress themselves are the easiest to use for those who can run the program. A major drawback, however, is that the executable file can typically be run on only one type of operating system (DOS, OS/2, etc.) -- thus excluding use by people on other systems. The singular operating system is determined at the time the executable file is created and will usually be the same as that on which the file is created. Versions of both *pkzip* and *pkunzip* are available for most operating systems.

If paper documents are currently produced, the contents of these documents can be represented in a ASCII text file. However, since ASCII files are not always a mirrored representation of the original, information on font definitions, special characters, artwork, proportional spacing, and other page layout characteristics may not be reflected. It is not uncommon that many of the characteristics of highly formatted files will not automatically translate to ASCII. If it is critical that the ASCII file maintain format characteristics, it may be necessary to embed codes -- or "mark-up language" -- in the file, to assure that these traits are not lost.

The ASCII mark-up language simply instructs software to display the file showing special characters and other non-ASCII features. This language enables users to view or print a document, with all its formatting, as it would be seen if printed directly on paper. Among the

more prominent languages that do this are two from *Adobe* -- *Postscript* and *Acrobat*. Although both of these processes more closely approximate the characteristics of the printed document, there are several disadvantages of such additional formatting information:

- file size increases,
- users are limited to those who have access to the software necessary for display or print, and
- data extraction is more complex.

It generally requires a skilled programmer or an avid user aided by good software tools to reformat data from an ASCII replica of a printed document. If the data being disseminated are likely to be imported by software for analysis, the data producer will probably want to create a file designed explicitly for that purpose. One way to do that is to provide the data in either a spreadsheet or a database. This procedure makes it easy for users who have access to the software designed to read those files, but relatively difficult for people who use different packages. Some files -- such as ones for dBASE -- are ASCII-based and may be reformatted relatively easily for use by other files. Some software products will sensibly import ASCII files if they meet certain specifications. For example, *Lotus 1-2-3*, *Quattro Pro*, or *Microsoft Excel* will import ASCII files, treating numbers as numeric cells and text as text cells, if the text fields are enclosed in quotation marks.

Binary Files

Binary files can be thought of as including at least two distinct file types:

- files that store information in a binary format (0 or 1) and require special software to extract the data, and
- files that are actually programs that provide data when they are executed.

These file types were touched on briefly in the previous section, when discussing compression of ASCII files -- one explodes the compressed file, either by using special software or by executing the file the same as any other compiled programs; others require specialized software to be read, like *Lotus 1-2-3* spreadsheets or files created by word processing packages such as *WordPerfect*, *Word for Windows*, *WordStar*, etc. With the emergence of the new multimedia technology, examples of effectively storing data as an executable file abound; for example, encyclopedias and other reference books have become available on compact discs and are read-accessible only through software stored on the CD-ROM disc.

Storing data in a binary file that must be accessed through specialized software offers the advantage of providing a readily usable product to those customers already using this software; in contrast, it makes it difficult for users who do not have access to software that reads those files. Agencies choosing such an approach should recognize the software constraints this places on the user and make their decision based on knowledge of the customer base. Another approach would be to provide a data file which may be read by several software systems. For example, a *Lotus 1-2-3* file is a relatively accessible binary file that many

different spreadsheets can import; it also can be read by users with different operating systems. Many people who analyze microdata use SAS, and a SAS transport file can be read by SAS on any operating system. Even so, providing data in only one format may restrict the audience for the dataset.

Storing data as a program allows the producer to provide easier access to the data for the user, but requires more effort on the provider's part. A well thought out program that presents statistical data could have any number of features that would make it easier for the user to extract the desired data:

- a user interface that is intuitive to a novice;
- the ability to extract data in several different formats;
- the ability to select only a subset of the dataset for extraction;
- painless access to documentation on the data being supplied; and
- analysis tools.

All of these enhancements, of course, require considerable effort by the data provider. It is the same as publishing software; programmer time is needed to put together a nice package. It will also require more testing than a normal data distribution, in order to prevent the release of programs which contain "bugs." Furthermore, if an agency wants to maximize the availability of its data, the programs will need to be ported so they can run on any of the several operating systems (DOS, MacIntosh, OS/2, Unix, etc.) in use today. Elaborate programs may also limit potential users by requiring them to have fairly high-powered hardware.

If sophisticated encryption is used, storing data as a program might also solve some of the problems involving confidentiality and disclosure.

Documenting the System

Thorough evaluation of technological alternatives should include consideration of the potential roles of hardware, software, and telecommunications components of the overall system. Having decided upon the type of system that will be implemented, both system access and use need careful documentation. The system design should also provide adequate technical support. Agencies should offer technical support, both as on-line electronic help -- for human/computer-interface oriented users -- and supplemental support by telephone. Important aspects of documenting any electronic access system are the specifications for system initialization and system navigation (i.e., loading and searching or browsing). More crucial than either of these tasks, however, is the written material that describes the system. There are two kinds of documentation that must be made available:

- material that describe the dissemination system and its use and
 - documentation that describes the data.
-

Organizations that currently disseminate data in hardcopy are already familiar with preparing dataset documentation. Today's challenge is to transfer these well-documented processes to an electronic system and to improve them with the newly available technology. Much of our discussion will focus on techniques to facilitate this transfer. Documenting the dataset for hardcopy purposes involves, at a minimum, specifications for file format; variable names, lengths, and attributes; and the typical record layout. Such documentation normally includes definitions of terms, statistical weighting procedures, copies of data collection instruments, as well as instrument reliability measures. It is also not uncommon to find that this document contains a detailed description of data collection procedures, variance derivation methods, estimation procedures, and sample design specifications. Frequently, such documents are used to provide information about restrictions on data use and interpretation. Although many users may not be interested in this level of documentation, organizations are encouraged to ensure such information is packaged with their electronically disseminated data.

Getting Started

The support issue always starts with answering the question, How do I get started? The response to the question is often dictated by the system that has been designed and by the mode of electronic dissemination. The entry point for any system will always begin with a written communication, even as minimal as a telephone number. For the typical telephone system, the user begins by dialing in a specified number and receiving recorded instructions. Upon establishing a connection to a modem or dial-in type system, users are generally provided several screens of supplemental information about navigating the system. In many instances, information and data are packaged for distribution, along with a set of written instructions on getting started. Frequently, in the case of data presented on diskettes, CD-ROM or other optical media, the distributor affixes a printed label onto the media, providing the basic information needed to "initialize" the system. Such instructions are usually brief commands, such as "at the DOS prompt, type A:\SETUP," etc. Other information systems provide more explicit instructions.

In other words, regardless of whether agencies decide on recorded media or on-line electronic transmissions, provisions for the basic start-up instructions are needed first, so users will know how to access or proceed with the use of the system.

Navigating the System

Once signed on, the next critical step is knowing how to navigate the system. A clearly defined method for proceeding is needed. Generally, this entails an extensive menu system and/or point and click techniques, utilizing either a graphical user interface with user-recognizable icons or hypertext screen paging. In recent years many user applications have been designed as *Windows*-based applications. The use of *Windows* technology provides users with a familiar environment. *Windows*' easy to follow standards for file management (print, open, save, and close), drop down menus, and editing tools facilitate users ability to readily understand how the system operates. Where *Windows* standards do not apply, consider using

menu systems, which walk users through systems one step at a time. A carefully designed menu system will reduce the amount and type of written documentation required to expedite uncomplicated access to information. In the absence of a menu system, it is very important that the graphical user interface be explicitly clear.

Another alternative is *Hypertext*, which permits users to jump seamlessly from one location to another. These links can be used as a means for associating various pieces of on-line information. A link could be established to associate a table of estimates or a cell within the table to text-based information. The text could be the value definition or it could be the statistical algorithm used in deriving the value. *Hypertext* is also a convenient method for associating one text segment with another. Clicking on a section title in an index, for example, could result in jumping to the section or chapter in which that title is detailed. A properly used *Hypertext* approach will also move users backward through their progression, so that they will always be able to backtrack to the point of entry.

Independent of whether a menu system or *Hypertext* approach or a combination of the two is used in the data dissemination system, an automated electronic retrieval system would be considered incomplete without some type of on-line help module. This help system could be provided either as text, similar to the **Help** feature provided by most *Windows* software, or it could be simply an extensive series of prompts to guide the user through the system. Often prompts and other on-line schemes are not enough. It is virtually impossible, if not impractical, for an agency to address every possible situation that may arise. Therefore, the need for human technical support remains.

Technical Support

Providing adequate and appropriate technical support for electronically distributed data has become a growing concern to agencies responsible for collecting data and deriving estimates. After years of effort to establish and refine policies relating to release of data, these agencies are finding that many of their policies are either inadequate to address issues of electronic data release or simply do not apply. Years of experience with data users have provided agencies with the necessary background to draft policies that reduced the likelihood that their data may be misused and/or misinterpreted by customers. In the electronic dissemination environment, however, the user community is greatly expanded; hence, the user profile and behavior history is no longer relevant. As a result, agency administrators have, once again, become concerned about misuse of their data. Thus, "field tests" of the data user community are needed to gauge the usefulness and the effectiveness of the proposed dissemination approach and to pinpoint where more documentation is needed. Furthermore, electronic data users need to have easy access to the documentation on the data, as well as the delivery system.

The rest of this section touches on three important areas of technical support for any new data dissemination system -- metadata, customer service support and data security.

Metadata.-- Despite built-in Help components, there will surely be users who need more assistance to understand or use the system. Mechanisms for support of use and interpretation

of statistical data have long been established. Previously, information systems were paper-based and support methods began with published documents that addressed the many aspects involved in deriving final measures. These aspects included, in some cases, a detailed description of study design and scope, survey methodology, along with general sampling and estimation procedures. It has also become an expected norm that this documentation would be supplemented with data editing and file creation procedures. The underlying rule of thumb is that enough information should be available so that replication of any measure generated is possible. Usually these reports are written to appeal to two groups of users -- those with a general interest in the subject matter and those whose interest is research-based. These reports should provide technical content that is of value to research methodologists, yet, will be informative to wider audiences.

In order to help prevent misuse of the data, users need to understand the survey concepts, as well as the limitations of the files. To encourage increased understanding about the data, electronic publications need to deliver this information quickly and conveniently. Fortunately, electronic media provide a unique opportunity to incorporate these kinds of data into public-use files -- to help alleviate misinterpretation of the data -- and agencies are urged to build statistical metadata components into their data dissemination systems -- including data and information about the data, such as reliability information and the description of the survey methodology.

While space limitations in printed publications often relegated standard errors to the back of the book, standard error estimates can be easily accessed by users of electronic publications. Where generalized variance functions are used, algorithms can be programmed to compute standard error estimates for any data item in the electronic publication. This is similar to practices in most agencies that permit easy calculation of standard errors from public-use microdata files.

The World Wide Web provides the opportunity to quickly access related information in different reports. Links can be created between statistical reports and survey methodology reports. Similarly, statistical reports that use data from several surveys can be linked, not only to more detailed statistical data from these surveys, but to the survey methodology and to the information on the data limitations. Statistical reports from one agency can even directly link to methodology information provided in another agency's report.

Using a *Hypertext*-type approach, data users may soon be able to instantly access all of the metadata associated with any estimate. The target and sampled population, the exact wording of the questions on the survey instrument and associated definitions, the standard errors, the item nonresponse rates and imputation methods, and other limitations of the data would all be available at our finger tips (or voice commands) for any estimate.

Of course, electronic dissemination of metadata has resource implications, as well. The more effort put into providing easy access to survey concepts and data limitations in the electronic publications, the less dependent users will be on the human resources at the agencies. Also, easy access to the metadata should decrease the likelihood that the data will be

misused. Furthermore, linking such reports to each other lessens the need to duplicate the same information -- such as the description of the survey methodology -- in subsequent reports that may be produced from that survey. One thorough methodology report may suffice, rather than adding an abbreviated methodology section in each of the reports in that series.

As new methods of electronic dissemination become available, the opportunities abound to better serve users and to more effectively use resources. Improving the dissemination of the metadata is an integral part of that process.

Customer Service Support.--Personal contacts to assist in the use and interpretation of statistical public-use data have always been available and have become increasingly important. Agencies provide this support by way of subject matter seminars and lectures by leaders in the field and via telephone access to subject matter experts.

In an electronic dissemination environment, this kind of interpretation and use support is especially important and, hence, more costly. However, careful use of on-line electronic methods may reduce the resource requirements for as many person-to-person exchanges. With this interest in mind, it is wise to incorporate both on-line and human interaction support into plans for the electronic system. In objectively designing an electronic system, beware that:

- all users will not interpret a given instruction in the same manner;
- a system's internal support mechanisms may fail to provide the level of detail desired by some users; and
- the total support structure may not be responsive to users with inappropriately configured hardware.

Therefore, expect the unexpected! In reality, with appropriate resources and an organized approach, user support issues are not insurmountable. Problems associated with user support for electronic data dissemination should not be considered unique. In fact, they are quite similar to those faced by software distributors and, as such, can be resolved using the same tactics.

The basic philosophy on user support for electronically distributed information is that each electronic data repository or interactive system should have:

- some type of hardcopy or on-line documentation that explains how it operates;
- a hardcopy reference manual or on-line Help feature to provide terminology definitions and instruction on how to perform specific tasks; and
- a means for contacting a real person, if all else fails.

Provide a telephone number for user support. Specify the hours of operation. In many instances, support organizations provide answering devices with a recorded message that contains the business hours, so that individuals who phone during non-office hours are notified of

appropriate call-back times. Finally, an option offering e-mail contact should also be provided. (This is a good way to get customer feedback on what is working and what is not.) You might even set this up so that some basic questions are always asked when someone “calls” in -- the responses would give you more insight into your customers, so that you can better tailor the data to meet their needs.

Privacy, Confidentiality, and Data Security.-- An equally important consideration to agencies is the assurance that confidentiality is not compromised. This is both a privacy issue and a data security issue. New modes of electronic dissemination may enlarge the user community for sensitive data, so that it includes users not considered under traditional security agreements.

Confidentiality assurance measures and data security procedures have both been in place for many years and must remain common practice. Barriers and password protection methods are critical to reduce the likelihood of an intruder. **Rest assured, however, that no interconnected electronic system is really impenetrable.** Traditional approaches to providing confidentiality, particularly among agencies that use human subjects as the basis of much of their published statistics, are well known. (See the FCSM's two Statistical Policy Working Papers on statistical disclosure.) Many of these same tactics will also apply to electronic distribution systems. Census takers, enumerators, interviewers, data and document handlers at the agencies should all be required to take an oath of confidentiality. Data analysts and computer programmers must ensure that identifiable information is not accessible to any unauthorized party, either by direct or inferential methods. Electronic distribution of data, particularly in systems where data are accessed using on-line methods, are challenged with the burden of prohibiting security breaches.

Data integrity cannot be ensured with the current state of computer security technology. Agencies must not only take care in the preparation of the data, but must also ensure the presentation of its correctness after it is placed in the public domain. A firewall is needed between the data to be distributed and the data construction process. This could include the use of separate computer systems for the production of the data and for the dissemination of the data. It is important to have mechanisms that allow users of electronically disseminated data to establish that the data they are using did, indeed, come from the claimed source and that no modifications have been made to it.

3: Operation of Electronic Dissemination Service

Preparation for Product Release

As detailed in Chapter 2, even if the physical infrastructure is in place for electronic data dissemination, considerable preparation must still be made for data or information release. For example, for an on-line system, a computer with dial-in capabilities or bulletin board software must be established; media replication facilities must be arranged for dissemination of recorded electronic data. The technical documentation and user support capabilities must also be in place before data release. While Chapter 2 laid out key issues for selecting media for a dissemination system, this Chapter will focus on the operational process.

Many people assume that once the system is in place, data or information releases can successfully occur with little or no planning. The prevailing misconception is "build it and they will come." The reality is that nearly every successful public release of electronic data should be planned in advance. Technical documentation needs to be prepared way ahead of time and user support operations need to be informed -- and possibly trained -- to handle questions about the new information release. Any data product inventory (for order handling) must be updated and made available to support personnel in advance of the release. In many cases, a number of technical decisions may also be needed before the data become available -- How big will the release be? How will it be packaged and presented? Will the typical customer be able to handle it in the format and size planned? For example, will users have to buy special software and computers to handle it? Will the customer understand information in the release without additional supporting materials?

Other planning must be done concerning release policy, as well as marketing of the information product. Will all data that are available within the agency be released to the public? How will the release be timed? If the public has direct access to government computers, how will security "firewalls" be implemented, so that they do not have access to confidential or business sensitive data, or to information related to "national security?" How will customers learn about the release? How will the release be priced and is this consistent with other information products currently in the sales inventory? The sections below will attempt to elaborate on some of these key issues.

Types of Data and Information Releases

Although all of the tasks or functions noted above must be performed regardless of the product, their complexity will differ depending on the type of data or information product. Complexity of preparation also differs depending on whether the product has a scheduled release, an unscheduled release, or an embargoed release.

Scheduled Releases

A scheduled release usually is one that comes from a large on-going information or statistical program and is planned as one of the regular "publications" of that program. They may be tapes, CD-ROMs, or on-line data files. They are typically the easiest release to handle, except that the electronic products are frequently very large and require advance planning with regard to policy, pricing, and user accessibility considerations.

Unscheduled Releases

Unscheduled releases are data or information products which typically have created their own demand and are being released due to public request. Advance planning should try to accommodate a certain number of these, although they can be difficult to handle. Documentation and/or user accessibility may suffer due to the short time-frame usually associated with these products. They also frequently become the orphan children of the sales inventory, because of poor documentation and inadequate instruction of the support staff. After the initial release and flush of publicity, later inquirers may find that the support staff has "forgotten" about the product's existence -- if the product even still exists at all, to say nothing of the accompanying documentation.

Two factors to consider in these types of releases are:

- Assuming the release is for publicly-available data, all members of the public should have equal access to the information. If the product is not incorporated into the sales inventory and the support staff is not informed of the release, then not all potential users will be able to acquire the information product. This can backfire if it is interpreted as favoritism and may subject the agency to legal action from those potential users who were excluded.
- If unusual technical "conventions" are used for the release media, many potential customers will be excluded who do not have the special capability. This can also be interpreted as favoritism and may cause problems for the agency later on.

Embargoed Releases

An embargoed release is one in which the data product is planned and prepared in advance, but cannot be released until some pre-specified time. Usually the release date (and time) is established in advance, and the information is prepared for release in advance, but is placed under "lock-up" until the release date and time. Many of the releases on the Department of Commerce's Economic Bulletin Board are embargoed releases of economic indicators. Such time restrictions are necessary to ensure that no one person or group has prior information which could give them unfair economic advantage.

An important policy question to be decided during planning for a new data dissemination system is where the information will reside before the release. Many originating organizations will not let disseminating components obtain the information prior to the release time. This usually means a delay while the information is brought "into" the dissemination system.

The trade-off for having the information in advance (for the disseminating organization) is the liability of possible premature "leaks" of business-sensitive, confidential, or government "secret" information.

Despite the timing considerations, it is important that embargoed information be released with the same criteria as other types of releases -- so that all citizens have equal access to the information. An embargoed release on an on-line system with inadequate phone access may, in fact, be more likely to subject the agency to accusations of favoritism and incompetence, because the data are so time-sensitive.

Documenting the Delivery System

As noted above, it is important to produce documentation for all electronic products in advance of their release. This is especially critical if the staff is new to electronic dissemination. Since most statisticians and information specialists are most familiar with paper information products, they may fail to realize how important this difference is. Many paper products do not need explanatory documentation or are self-documenting. Remember, though, that even the release of text information files requires directions on how to use an on-line system or a fax system. With rare exception, this part of the process is ignored; consequently many electronic products released by the government are characterized by poor documentation.

Announcement of Releases

How the announcement of information releases is handled varies greatly from agency to agency. In years past, not many formal mechanisms existed for the announcement of agency releases. The Office of Management and Budget now produces each year the *Release Dates for Principal Federal Economic Indicators*, a comprehensive calendar of releases of economic data. The *Federal Register* and *Commerce Business Daily* also provide mechanisms for officially notifying the public of data releases.

With the advent of programs like the National Performance Review, more agencies today feel the need to go beyond these official avenues to try to reach as many members of the public as possible. Agencies are much more involved in "announcement" activities that resemble private industry marketing exercises. These include mailing flyers to targeted mailing lists, compiling agency catalogues and newsletters that document product releases, and even some degree of advertising in printed publications -- including some paid advertising. Not surprisingly, recent trends indicate that more and more agencies are also turning to electronic sources -- like bulletin boards and the Internet -- to make these kinds of announcements, as well. Whatever the means, in the current climate a proactive approach to announcing information products is certainly recommended.

Order Processing

When an information or data product has been made available to the public, plans are also needed for handling the requests that come in. Usually this involves an order-taking capabil-

ity, which forwards the data requests on to areas which fill the orders. Rush orders, replacement orders, and management information on the number and type of customers are all considerations involved in establishing an order-handling capability. Most organizations try to create an automated customer database which will handle this function, where phone staff and clerks input the orders into the database.

If there is a user fee or charge for the information, the complexity of the order handling is greatly increased, and all of the accounting problems and rules associated with handling money become criteria. It is important to remember that customers are much more serious about the orders if they have paid for them and are less tolerant of delayed or lost data requests. Because of the need to carefully track all funds received by Federal agencies, charging and receiving money for information products is not a task to be undertaken in a cavalier manner -- in fact, some agencies will opt for a free distribution for this reason.

Even if the information products are not "free," inevitably there will be a large free distribution for political or public relations reasons. In setting up the ordering system, criteria and procedures need to be established in advance for handling these free orders in an efficient and equitable manner.

Order Fulfillment

Next is the process of "getting" the electronic information into the hands of the customer. This section looks separately at the two major categories for electronic data products in terms of order fulfillment -- on-line electronic transmissions, for which data communications links are used, and recorded physically-transportable media, which include all kinds of "other media" ranging from CD-ROMs to computer tapes with print-outs.

On-line Transmissions

On-line products are characterized by the ability of the data customer to control the timing of dissemination once the product has been released. The prevailing misconception with on-line products is that once the systems are in place that they "run themselves." In reality, staff must maintain and update on-line systems on a fairly regular basis. One of the main strengths of on-line systems is that new products can be made available through them very quickly and simultaneously to all users. To capitalize on this strength, planning and execution for on-line dissemination should be performed concurrently with planning and execution for other dissemination efforts.

Careful attention should be given early in a project to various technical issues which can affect the scope of the on-line effort. How does the on-line product complement or overlap with the other electronic and non-electronic products? What are the hardware and software requirements for the planned dissemination? What level of detail will the on-line system provide? Will traditional products be redesigned to take advantage of the strengths of on-line systems?

While the area of electronic transmissions is rapidly changing, current examples of on-line systems include "fax-back" servers, "dial-up" bulletin boards, and "Internet servers." As Chapter 2 details,

- Fax-back servers allow the user to select publications or documents on a phone menu system and have copies of the selected items sent back via fax transmission. Some support personnel may be needed for this effort.
- Dial-up bulletin boards permit the user to connect to the provider's computer via modem. One advantage is that usually the customer is given the capability to leave suggestions or messages for the system operator. This capability requires someone to review and possibly reply to those messages at reasonable intervals.
- Internet servers let the user connect to the provider's computer via the Internet. Current applications, which are already widely available, include the World Wide Web, gopher, archie, ftp, and electronic mail servers, as well as Internet applications which may become widely available in the future. Internet servers require a connection to the Internet for both the user and the product provider.

One of the important issues of particular concern for product providers using the Internet for dissemination is network security. In many cases, to protect their own internal networks and data from outside intrusion, product providers will need to insulate their Internet connections from the rest of their network. Many persons, when planning for initial Internet connections, fail to recognize that there may already be security policies and procedures internal to their organization which apply to their planned activities. Maintaining a proper security environment is an ongoing process, which requires the staff be aware of new developments in the rapidly changing field of computer networking and to constantly monitor the system for potential intruders.

Perhaps the most important activity for ongoing maintenance of an on-line system is to schedule regular tests to verify that the system is actually functioning as expected. Nothing is more frustrating to a user than to find that on-line systems are not, in fact, "on-line." On-line systems require continual maintenance of links to ensure users can periodically return to previously used addresses and can determine if new data or revised data are available. Clear lines of authority and responsibility should be established from the outset for making sure that the electronic transmissions are working properly and that means are in place for correcting problems when they develop. Smoothly running systems require allocation of responsibility for such functions as hardware repair, software maintenance, and regular backup.

Finally, as mentioned earlier, a main strength of on-line systems is that additions and updates to the available products can be made easily. Many providers will desire to establish protocols through which these changes can be approved in a timely and orderly manner. To streamline that activity, the clearance process for release of on-line products should be packaged as part of the usual process for other products. The main difference between on-line

products and more traditional publications in this area is that, in most cases, the data can be released virtually as soon as the approval process is completed. For example, on-line products do not require time for printing, as publications do; if the clearance process for publications is structured appropriately, release of electronic publications can precede or coincide with their release in printed form.

Recorded Media

There is an essential difference between on-line or electronic transmission products and recorded media products in terms of order fulfillment responsibilities. The on-line products involve data communications between the supplier agency and the user, while recorded media are physical electronic products, like computer tapes, diskettes, and CD-ROMs. Whereas the customer can directly establish a data communications link with the agency to receive on-line products, in order to obtain recorded media products, it is usually the responsibility of the agency to "ship" the product to the user. The user, then, may or may not need special facilities in order to access the data -- e.g., specialized software or hardware for CD-ROM. There are several ways in which these physically-transportable products can be distributed -- as scheduled products, single orders, or rush orders. Procedures will be needed to address each of these options.

Scheduled Products.-- Scheduled products are those released and handled on a prearranged basis -- for example, by subscriptions. In these arrangements, the user buys a certain number of issues in advance, and prepays for them. Many times, these issues of the product are released and shipped on a pre-established time schedule. For example, certain economic or business data might be released and shipped at a certain time of the month.

Scheduled products present many more problems for recorded media release than for on-line release. Arrangements to produce and ship the product must be made ahead of time and must adhere to a tight schedule. Any problems with the production process telescope into a short time-frame at release time, and these eventualities must be anticipated. Electronic media produced in "real time" are much more difficult to schedule and ship on time than manufactured products, such as CD-ROMs. Many a release schedule has been devastated by the slow through-put of tape drives on the "mainframe."

Single Orders.-- Order fulfillment for singles orders is much easier to handle, since customers will not expect these products on a pre-arranged schedule. There are usually customer service commitments which the agency will adopt, and the production and shipping process will need to be constantly monitored for adherence to those standards. Customer service standards, however, can be adjusted to reflect changes in the production and shipping environment.

Agencies will sometimes release products in a combination arrangement, in which the agency agrees to a "subscription" type of release, but shipments are made only as new, pre-specified products become available. The demands of this type of release are not as severe as a scheduled release, but users will complain if some tentative schedule is not met.

Rush Orders.-- In the current world of global overnight shipping services, the "rush" order is a reality for which the agency must prepare. Special care should be given to planning how products are shipped, what circumstances or customers merit accelerated service, and, especially, how the shipper will get closure on the delivery of an order. If this is not done, an inordinate amount of time will be spent by the customer services staff "tracking" orders and re-sending those that disappear. Beware that inadequate record-keeping can lead to duplicate orders and a subsequent loss of staff time and revenue.

Overnight services can be very expensive, and must be factored into the pricing and budgeting of the products. Foreign orders are another particularly "dangerous" area, because of their expensive shipping costs; policies for dealing with overseas orders should be established with the initial offering of the product. Finally, another uncertainty arises from the vagaries of the government bidding process -- agencies should plan for the eventuality of "losing" their overnight carrier contract -- contingency plans are needed for how to deal with customer expectations if this happens.

Customer Support

Another key issue which must be considered in the earliest planning stages of an agency's data dissemination process is customer support. Decisions about appropriate data products must take into account the agency's ability to support these products and the customer's ability to use them. Major factors to consider include:

- What staff resources will be allocated to the electronic data dissemination program?
- Will the same staff responsible for traditional data dissemination services be charged with supporting an electronic data dissemination program?

Adding this new responsibility to existing staff requires adequate training in the use of electronic media. Decisions must be made as to the level of user support that can realistically be provided by the dissemination staff. For example, how will questions about use of software be addressed? Will users be able to electronically send queries to an agency's Internet mailbox? If so, will these requests be placed in the queue with requests received through traditional methods -- phone calls, letters -- or will they get higher priority?

Also, it should not be assumed that electronic release of data will lessen interest in traditional forms of data release. In fact, data dissemination programs should expect an overall increase in information requests, as new data users discover agency resources.

- Are the needs of both internal and external agency customers being considered?

Dissemination staff must work with **external** customers, to identify products to release electronically and to determine the best electronic format for that release. This should be based on assessment of customer capabilities and needs, including access to CD-ROM drives

and other hardware considerations (computer size, speed, etc.). It should **not** be assumed that customers know how to work with electronic media. Agencies now providing access through Internet, for example, are learning that many users need assistance simply to **use the Internet**. Customers may even need documentation to help them use electronic media and analyze agency data.

The staff directly involved in information dissemination must also work with **internal** "customers" -- the agency's data producers -- to determine which products to release electronically and the best format (CD-ROM, diskette, Internet, etc.) for release. Should entire files or only selected pre-tabulated data be made available electronically? Should this be released as flat ASCII files or with customized software for access and/or analysis? Internal customers also need feedback from the dissemination staff on customer satisfaction with data products and data release formats, to help guide future products.

Order Fulfillment and Accounting Support

Agencies will also need to determine appropriate dissemination outlets for electronic products. In addition to internal agency mechanisms, the Government Printing Office (GPO) and the National Technical Information Service (NTIS) both can handle distribution of electronic products. Using multiple vendors can increase access to data and possibly lower dissemination costs.

Updates, Corrections, and Errata

One of the strengths of an electronic dissemination system is the ability for on-line notification of updates, corrections, etc. But customers need to know where to look for this information. For example, agencies may have an icon or menu selection on their Internet server to identify new products or to announce new releases on a bulletin board of their own or through the NTIS or GPO bulletin boards. Agencies should also have a method for users to electronically query them about updates and corrections -- i.e., with an Internet mailbox.

Archiving

The word "archiving" can have two meanings for managers of electronic dissemination programs. In its best known context, "archiving" is synonymous with "back-ups." It means the on-site, temporary storage of critical information records, in case of system failure.

Another aspect of archiving is the permanent retention of information "records" for posterity, accomplished for Federal agencies by planning and implementing a records acquisition schedule with the National Archive and Records Administration.

Back-ups

Archiving on-line and recorded media requires an awareness of general archiving issues, paired with an appreciation of the specific difficulties in archiving electronic data. Most

organizations already have guidelines relating to when, how, and where to archive their data products. These guidelines may need to be modified -- or media-specific guidelines may need to be developed -- to handle electronic media.

Almost all electronic media are less durable than traditional printed material. Some -- such as magnetic tape -- require much more careful handling and storage under more controlled environmental conditions than printed material. Many organizations already have existing capability and expertise for storing electronic media, particularly in mainframe computer environments; often these can be tapped to provide services for the dissemination effort. If such facilities are not available, outside services may be needed. However, particularly with the advent of diskettes and CD-ROMs, archiving is now possible in a personal computer environment at relatively low cost.

Regardless of the media used, special care is required for archiving electronic products and documentation, to ensure that they appear on the media in a usable form. Especially for long-term archival purposes, unusual or obscure data or file structures are obviously not recommended. Remember -- changes in the computer industry happen so quickly that today's de facto **standard** format may become tomorrow's dinosaur!

Problems with data formats have historically been particularly acute in the personal computer environment, but have also been apparent in mainframe environments. For example, be careful that you don't archive the data without preserving the software to read the data file -- especially where microdata are involved.

Because of the problems that can arise with improperly archived data, it is prudent to consider saving at least crucial files in more than one format -- especially if no practical NARA standard exists for the agency's electronic products. Saving a second copy of products in a simple format that can be read by many different software packages -- such as open ASCII text, for personal computers -- is one option. Another possibility is to keep track of formatting changes in the computer industry and to update formats for key data as the "standard" formats change. That is probably a good idea, in any case, where important historical data are involved. In any case, it is advisable to avoid complicated or proprietary formats entirely. Unfortunately, this may make your product less desirable and useful for more sophisticated customers. Even printed versions which document government data and systems are useful to maintain in the archive, so that users can reconstruct the files, if necessary.

NARA Activities

The planning for electronic dissemination should also include the last stage of the information life cycle, namely, the final disposition of the information -- either by destroying the records or retiring them to the National Archives and Records Administration. This is a formalized process in which NARA is contacted and the information is evaluated by the agency and NARA as to whether the information is suitable for the permanent archive. If it is, NARA will schedule it for acquisition and the implementation of this schedule involves using official procedures, media, and standards, as established by NARA. The costs of providing the information to NARA are undertaken at the agency's expense.

Because of the expense of this step, some Federal managers ignore or avoid this requirement. It is critical to view it as an important step in the information life cycle of an electronic information product in a successful dissemination program, and (if properly done) can eliminate some agency expenses associated with saving dated versions of the information that the public may request. It can, in effect, eliminate the need for extensive "back-ups" of critical information over long periods of time.

4: Customer Service Programs

Introduction

Although customer service programs are not new, they have become a key component of the Federal government's program to maintain quality management and implement high standards of customer satisfaction. This chapter describes several key components of an effective customer service program for any agency, especially one engaged in the electronic dissemination of statistical information.

One word of caution -- developing customer service programs is an expensive and painstaking task and maintaining them is a permanent commitment of resources. Even in tight budget times, the expectations they raise and the good will they engender are difficult to ignore or roll back. Therefore, establishing these programs should be closely tied to the agency's strategic planning process, so that an overall program mission and specific program goals and objectives can be developed. Then, the agency can establish mechanisms to capture customer feedback and select performance measures to monitor the program's progress toward achieving its goals.

This is especially important for agencies embarking on electronic dissemination programs. Agencies should anticipate that the future will bring improvements in technology, as well as raising the level of customer sophistication with respect to electronic dissemination products. This will cause customers to become increasingly more demanding regarding the level of service provided. The level of service which is satisfactory at the outset may be less than satisfactory within a short period of time. Thus, through customer feedback, the established evaluative criteria must be continually modified to reflect these rising expectations and balance them against available resources.

Setting Up a Customer Service Program

Many agencies have established customer service or information centers as focal points within the agency for assisting users. The staff locates agency information, provides the information in multiple forms, offers technical assistance and support, as well as making referrals to the appropriate specialists or other sources, if needed. Agencies that choose not to establish a customer service center will need to train staff throughout the organization to deal with technical assistance and support, as well as questions related to the availability and interpretation of the data. This section focuses on the activities and services that a customer service center should offer to support its electronic products. Similar activities and services should be offered in a decentralized customer service operation.

Basic Activities of an Information Center

Each agency should maintain a staff or center of trained information specialists available to respond to requests from a wide range of customers for the agency's statistical information and

information about the agency's statistical products and services. Requests may be by telephone or letter or by electronic means, like e-mail and Internet. The center may also provide automated telephone service, featuring recorded information or menu-access to specialized staff support.

Center staff should also be responsible for developing a series of materials designed to supplement information provided to individual users upon request. These materials should be developed with user input, to assure maximum value:

- directories of products and services;
- directories of data systems;
- directories of agency-produced software;
- latest scheduling changes or updates of data/software
- customer newsletters;
- directories of information sources;
- fact sheets and promotional materials;
- press releases; and
- order form flyers.

These materials should be used proactively to inform users of product and service availability. For instance, the order form flyer contains product description and ordering information and users can mail or fax it back to the information center. These, generally one-page, descriptions can be inserted in other mailings or reproduced in printed reports or elsewhere. Agencies should also consider distributing electronic versions of this information through their bulletin board or Internet services.

Concerning electronic products, each catalog or newsletter should contain the official name of any product, including any unique identifying codes, and a reasonable description of the product, including the major data items -- the time span covered by the data; whether the product contains summary or microdata; the format of the data (e.g., ASCII, dBase, WKS, etc.); whether the product comes with software or the type of software required; basic systems requirements; whether the electronic product corresponds to a particular printed product; and ordering information, including source and price, plus order forms.

This information, primarily in printed form, should be made available at no charge. Similar information or listings can be posted on-line, as well. In either medium, sample tables and graphs can also be provided, to give users a better idea of the contents of the product.

Assistance in Using Agency Products and Services

Unquestionably, information center staff should do more than simply act as information finders. First, they should be trained to assist customers who have no experience with agency information or who have general or unfocused questions. They should also be trained to answer customers' questions about the technical processes for using any of the agency's products or services. This is especially important when the customer cannot make an electronic

product or service perform to his or her expectations. Historically, such process questions were internal to an agency and were handled by a data processing "help center." Today, it may be unclear whether the customer is experiencing an information problem or a technical problem -- or is simply not sufficiently knowledgeable about a product or service to make it work or be able to navigate it successfully. Therefore, center staff should be equipped to discuss both content and process questions. System and product documentation, users guides, how-to manuals, and sufficient training in the products and services available are key ingredients in a successful assistance program.

Technical Support Service

A variety of information dissemination capabilities are recommended for information specialists to service customers. These include:

- electronic bulletin boards;
- e-mail, listservs, and other Internet services through the local organization, as well as easily accessible on-line services, such as *Prodigy*, *America On-Line*, and *CompuServe*;
- fax services, combining paper and communications to respond to individual customer requests or to send data or publications to customers on a proactive basis;
- fax-on-demand services that are initiated by the customer;
- electronic data dissemination capabilities, including electronic response or computer-readable media, such as diskette or CD-ROM; and
- on-line network/dial-up available databases.

These services should be easily accessible, easy to use, capable of measuring user access and recording user satisfaction. Information specialists should have access to all available services electronically, with the latest available and fastest equipment on hand, in order to keep pace with the rapid changes in information technology and the rising expectations of their customers.

Agency-based Locator/Finder Services

In addition to the directories mentioned above, agencies may establish other information locator-finder services for their customers. Two examples are described below -- participating in the Government Information Locator System (GILS) and providing information referral services.

The Government Information Locator System.-- The GILS effort is a government-wide program to help the public locate and access public information within government invento-

ries. GILS is a network of decentralized, agency-based information locators, set up by the Office of Management and Budget. Although agencies are required by OMB Circular A-130 (June 25, 1993) to create and maintain an inventory of their information systems and information dissemination products, GILS will standardize the process of identifying and describing public information resources. The public will be able to use GILS directly or through intermediaries -- such as libraries, private information providers, and academic institutions -- and through various media -- including Internet and modem. Federal statistical agencies are encouraged to participate in the development of their Departments' "GILS core," but should also develop their own GILS-compatible locator services, to offer their customers a greater level of detail and accessibility. (Additional information about GILS is available at the following Internet address: <http://www.usgs.gov/public/gils/>.)

Referral Services.--The traditional role of information centers has been to provide access to certain types of data. The first level of inquiry service is generally to provide customers with information and information products produced by the parent agency. The second level of service involves customers who are seeking related information, not produced by the parent agency. Under this circumstance, an information center may decline to serve the customer or may volunteer to refer the inquiry. Anecdotal evidence shows that agency information policies vary widely in this regard. Some centers generally assist customers in accessing the agency's information, while others go as far as readily suggesting comparable or complementary data available elsewhere in the information community, public and private. Some centers will even produce guides to all comparable information and services, regardless of source. These guides can be either internal documents, to assist the staff of the information center in handling inquiries, or external documents, for general distribution. (For example, the Energy Information Administration's Energy Information Directory, although used by information center staff as a day-to-day reference work, is also made available to the public.) While producing and distributing such a guide may seem to take customers away from the center, in the end, the guide can attract more customers to the center -- regular customers become used to having all of their needs met and come to use the "full service" center much as a research library or information finder.

Marketing Programs

In the world of electronic dissemination, rapidly changing technologies are increasing customer expectations for timeliness, accuracy, ease of access, and creative new services. Marketing programs provide key tools to identify new markets, seek out customers in those markets, as well as traditional customers, to understand their needs, and establish promotional programs to reach all audiences, new and old. Joint ventures are an innovative way of reaching out to new customers. This section highlights some marketing activities that are especially useful for customer service centers which disseminate electronic data. Depending on how deeply an agency commits to marketing, it may wish to consider establishing a full-time marketing function.

Customer Identification and Tracking

An agency needs to know something about the customers it serves. Who are they? What do they need from the agency? In what formats? How can the agency best serve their needs?

One of the hardest tasks an agency will face is determining who its customers are. Is a customer a person or an organization? Is a customer only someone who purchases from the agency or is it also the casual inquirer? Are potential customers as important to the agency as current users?

A second task is deciding who are the agency's principal customers. Conventional marketing wisdom is that 20 percent of customers account for 80 percent of sales. Does a student doing a term paper merit the same attention as a Congressional inquirer or a major think-tank organization's request? Or, more importantly, how can the agency service both within their context, given limited resources?

Third, what kind of information does an agency need to collect about its customers? Is a simple name and address sufficient? Would it help to know what products are being purchased? Should every transaction or communication be noted in a database?

Fourth, what safeguards must an agency observe in handling privacy information? What are the limitations of a customer information system? Who has access to it and for what purposes?

Identifying Customers.-- An agency needs to differentiate as much as possible its different classes of customers. There are several different frameworks for doing so -- by standard industrial classification code, by a classification based on the agency's experience, or by a classification based on other factors, such as the customer's role in the distribution process or volume of sales.

A common framework is the standard industrial classification (SIC) code -- a hierarchical scheme that is used to classify establishments by the kind of economic activity they carry out. The SIC is generally used by the list-management industry because of its fine level of detail. The name and address lists of customers can be coded by 1-, 2-, 3- or 4-digit SIC code, depending upon the level of detail required. Besides the *SIC Manual*, the agency would require some standard look-up products, such as *Dun's Business Locator*. Once the lists were coded, an agency could see whether the bulk of its customers are in academia, government, or a particular kind of business, such as statistical consulting.

Another workable approach is to use coding appropriate to the agency. Based on the agency's mission, experience, and knowledge of customers, some variation on the SIC system might be more appropriate. For example, an agency's customers may be concentrated in academia and the agency may need to make a distinction based on the academic department. The remaining customers could be grouped into large general categories.

Other possible coding schemes could be based on any of the following:

- volume of sales or inquiries;
- the relative importance of the customers to an agency -- for example, other Federal agencies, the media, Congress, the general public;
- the role of the customer in the overall distribution process -- e.g., reseller, library, end user.
- the customer's occupation or position in his/her organization -- e.g., president, CEO, researcher, project manager, librarian or information specialist.
- the customer's preferred media -- e.g., magnetic tapes, CD-ROMs, on-line systems, or diskettes.

Being able to differentiate among the major customer types allows an agency to better define its products to meet specific needs of specific types of customers.

Figure B shows how customers are classified at the Energy Information Administration. Another alternative, shown in Figure C, is used by the Bureau of the Census.

Figure B: Energy Information Administration Customer Typology
<ul style="list-style-type: none"> → EIA → Other DOE → Other Federal → State/Local → Library/School → Foreign Government → Congress/Staff → White House → Research/Academia

Figure C: Bureau of the Census Customer Typology
<ul style="list-style-type: none"> → Private Citizen → Government Sector → Academic/Research Sector → Business Sector → Public Service Sector → International <p>These six sectors are further broken down into 22 subgroups.</p>

Customer Information Systems.--In order to keep track of customer information, a customer service center should develop a customer information system. Besides just capturing names and addresses, the agency should determine what other information it needs to know about its customers. The amount and kind of information that should be maintained will vary according to budget, staff, other resources, and use of the system. Of course, an agency should determine beforehand what it wants the system to do, the kind of reports management requires, and the amount of resources available to maintain it. An agency should not collect information on customers just for the sake of collecting information; rather, information should only be kept that will be used for meeting customer needs.

Obtaining basic customer information is one of the easier tasks. Most offices maintain records of calls or files of inquiries; agencies that sell products maintain lists of purchasers. The U.S. Government Printing Office can provide an agency with a list of report purchasers or subscribers for the cost of doing the computer run.

In order to make the system useful, basic customer name and address information should be standardized. This will help data entry, list mergers, segmentation, and qualification for discounted bulk mail. Any customer classification codes should also be included, to enable the agency to segment the list.

Below are other useful features which a customer information system could include.

- An indication of the source of the name and a date should be added to the record to identify how long the customer has been in the system. The date is also useful for later purging inactive customers -- i.e., those who have not purchased or contacted the agency since a specific date.
- If the agency sells its own products, some information on purchasing history -- such as the name of the product -- would be useful to maintain. Sales amounts could be helpful in some instances, as well.
- If the agency does promotional mailings, it may want to track the number of times a customer has been sent a promotion and the number of times that particular customer has responded. Such an analysis would be useful to evaluate the promotion and find out more about the customer's needs.
- An agency may also want to track the number of contacts with a particular customer and any feedback the customer provided. However, such a feature would require a large interactive system or database.

Agencies have the option of developing their own systems, contracting out development, or purchasing off-the-shelf order-entry, customer contact software products. The type of system and its cost would depend on how each agency wanted to use the system.

Market Research and Customer Surveys

A customer information system tells the agency something about its current users, but it cannot stand alone. To serve customers effectively, the agency must understand its products and services market. Too often, agencies make products available without researching the market for the products. At other times, the product is not in a desired format or the data are not presented in appropriate size-classes, so that users can easily turn the data into useful information. Market research into customer needs is critical. There are several research techniques that have proven effective -- literature searches, focus groups, and customer surveys.

Literature Searches.-- Literature searches are a good place to start. There are a number of good sources of information on trends in the industry that may help an agency evaluate the usefulness of its products and services. Some of these include *American Demographics*, *The Wall Street Journal*, CD-ROM and multi-media catalogs and publications, *Sales and Marketing Management* magazine, and various library publications. Most agency libraries can do literature searches and retrieve the latest articles on particular topics of interest.

Focus Groups.-- Focus groups, consisting of known and potential users, intermediaries, and secondary distributors, can help an agency refine its products. Users are generally brought together for as little as a few hours to as long as a day, and are asked their reactions to the proposed product or activity. A trained facilitator leads the discussion, while others record the focus group's responses. This sort of activity can provide meaningful insight into the customers' specific needs and requirements.

Customer Surveys.-- Executive Order 12862 on Setting Customer Standards requires "continual reform of the Executive branch's management practices and operations to provide service to the public that matches or exceeds the best service available in the private sector." Among other activities, agencies are instructed to identify their customers, survey them regarding the kind and quality of services they want and their level of satisfaction with existing services, benchmark customer service performance against the best in business, post customer service standards, and measure results against them.

To assist Federal agencies in carrying out customer surveys, the Office of Management and Budget developed a *Resource Manual for Customer Surveys*, that was released in November 1993. Figure D, below, summarizes the key components recommended for customer surveys. Agencies are encouraged to review the manual before embarking on any customer survey. The following sections provide summary information from that report on how to conduct customer surveys.

Figure D: The Steps in a Customer Survey

1. Determine the scope of the survey measurement program
2. Identify the factors and characteristics that underlie customer expectations, needs, and satisfaction
3. Identify the target customer population for the survey (possibly intermediaries rather than end users)
4. Develop the sample frame of the target customers
5. Choose a data collection method best suited to your customer survey
6. Choose who will collect survey data from customers
7. Develop and pretest the survey questionnaire
8. Construct the statistical design of the sample of customers
9. Design procedures to achieve high response rates
10. Ensure quality while the survey data are being collected
11. Process the survey data and prepare them for analysis
12. Analyze the data, summarize the results, and present the findings

Setting Guidelines for Customer Surveys.-- The agency needs to determine what it wants to measure -- satisfaction, use, desirable improvements; current and future needs; who will carry out the activity; and how the results will be used. A committee or office should be charged with coordinating or managing the agency's customer surveys.

Other goals/guidelines should be to:

- minimize duplication of effort within the agency;
- avoid repeated contacts with the same customers;
- assure that the surveys are well thought out and will produce statistically valid results; and
- ensure that the information will be useful to the agency.

Here are some suggestions for ensuring a good customer survey:

- Contact the same customer no more than once a year.
- Use a sample of customers rather than a census of the entire population, unless that population is small. You can use a systematic sample -- such as every fourth customer, every tenth customer, etc. -- or target different groups through a stratified sample approach -- e.g., if news media or Congressional inquiries are more important, those customers could be selected at a higher rate than the academic or private sector users. Also, one stratum should include product type.
- Keep it short. Limit your survey to a few questions. A page is ideal; four pages is asking a lot on a voluntary survey.
- Include some follow-up for nonresponse. A mail-back survey will generally have a return rate of between 20 - 30 percent. Telephone calls to nonrespondents of a Census Bureau customer survey increased the rate to 58 percent. A combination of reminder notices and telephone calls netted the Bureau of Labor Statistics an impressive 88-percent response rate.
- Ask actionable questions, such as "How can we improve this product or service?" or "If you do not use this product, why not?" and provide the customer with an opportunity to explain.
- Include some standard questions on all surveys. This will enable the agency to measure some items across products and programs for a period of time.

There may also be other questions that are of particular interest to the specific agency.

- Leave some space for open-ended comments. Responses can often be enlightening and reveal aspects of the agency's product or service that the customer considers important, but that may have been overlooked.
-

- Develop a standard format to report customer survey results; this will help make sharing and understanding information across surveys easier.

Customer Survey Instrument.-- The most common method of surveying customers is through the mail. Telephone interviews are also common. Agencies may want to consider personal interviews, where more in-depth probing on particular topics may be required. Other options include:

- on-line surveys, such as through the Internet or a bulletin board system;
- fax surveys, whereby potential respondents are faxed a short survey and requested to fax back their responses; and
- questionnaires within the electronic product, itself, that users can print out and mail back or return electronically.

Don't forget that all surveys going to respondents outside the Federal government must go through a formal Office of Management and Budget clearance process. OMB has established special guidelines and processes for customer surveys that are not as lengthy or as involved as those for other surveys. The agency should review its plans with OMB's Office of Statistical Policy.

Rather than measure user response to a particular product, agencies may want to measure general satisfaction over an extended period of time. The following are some suggested activities in this area:

- a product evaluation postcard -- with return postage paid -- included with every product sale, asking general questions about customer product/service satisfaction;
- a more in-depth questionnaire, mailed to a sample of purchasers or inquirers;
- selected questions asked of all or of a sample of phone inquirers on a given day or during a given week;
- follow-up in-depth telephone surveys of inquirers who have agreed in advance and specified a time for call back; and
- a broad or ongoing focus group of users of major products, for continual feedback.

Whatever the mechanism, the same questions should be asked over a period of time, so that improvement can be measured. In fact, the agency may want to establish an overall rating, with a combination weighted score that can be monitored.

Promotional Activities

Advertising

Several standard promotional tools are available to agencies to inform the public about products and services. An agency should look closely into the costs and benefits associated with these activities before embarking on any promotional campaign.

Promotional tools include:

- paid advertising, including such media as newspapers, television, radio, direct mail, and magazines;
- unpaid advertising, such as professional newsletters, other agency publications, bulletin boards, and listservs;
- exhibits and conferences; and
- personal contacts.

Paid Advertising.-- In general, an agency may inform the public about its products and services through the use of paid advertising but should avoid promoting the importance of its programs and activities. (However, any agency wishing to buy advertising should first ascertain whether the legislation that governs its activities contains specific guidance or restrictions concerning advertising.)

Paid advertising can take several forms. Direct mail can be focused at a targeted audience or saturated within a geographic area. Advertisements in professional or trade journals also target specific audiences. Advertisements in major media reach a much wider audience than the other forms, but will likely have less payoff.

Unpaid Advertising.-- In the area of unpaid advertising, there are several channels available to agencies to take advantage of free exposure in newsletters and elsewhere.

- Articles in newspapers, generated from press releases, often result in awareness and, occasionally, sales. Be aware, however, that the media have been known to overstate or misinterpret data provided in press releases.
 - Many agencies cross-advertise in each others' publications.
 - Newsletters of other organizations can reach more targeted groups. Many state and local agencies and trade and professional associations have regular newsletters which could serve as additional promotional vehicles. The Association of Public Data Users, the Council of Professional Associations on Federal Statistics, and *American Demographics* magazine regularly report on new statistical products and services. Other potential sources of publicity include computer trade journals, library journals, and commercial product directories.
-

- Electronic bulletin boards are a newer source of advertising. Several Federal agencies maintain electronic bulletin boards that would welcome information about an agency's products.
- Internet offers the possibility of providing basic datasets, as well as stimulating interest in the agency, its products, and promoting product sales. Several businesses and some agencies offer order placement through Internet or provide promotional materials and order forms that can be downloaded and mailed, faxed, or distributed by listserv.

Exhibits and Conferences

Many Federal agencies participate in association meetings, trade shows and conventions, either as participants or exhibitors. Exhibits serve two basic purposes:

- general outreach for the organization, including as a recruiting vehicle, and
- exposure for new products among data users.

Before embarking on an exhibits program, consider the issues of equipment and cost. Specialized equipment is needed to present a professional appearance at a trade show or convention. Up-to-date displays, product samples, brochures and flyers, and, perhaps, videos or computerized demonstrations are among the items needed.

Obtaining these, however, is only part of the cost. Agencies also need to consider costs associated with maintenance, shipping, drayage, space rentals, and services and furnishings while on the exhibit floor, not to mention staff travel and per diem.

Personal Contacts

Finally, an agency should consider the marketing impact of speeches and workshops at conferences and conventions, and the good will and personal satisfaction that comes from personal contacts. Moreover, this may be the most cost-effective approach for some products, especially complex ones that require more sophisticated knowledge or technology to use properly and, therefore, benefit from exposure in a setting where staff are available for a "hands-on" demonstration.

Joint Ventures

The role of joint ventures in the Federal government is not well defined, as yet, and needs further investigation. In general, it is the sharing of profit -- as well as of risk -- in the marketing of a new product or service. The groups involved may be two Federal agencies or a Federal agency and a private sector group. Roles and responsibilities are generally specified in a formal written agreement.

Joint ventures may include "publishing" the agency's data, providing software, acting as a sales agent, etc.

Sellers, Resellers, and Redistributors

Some agencies have contracted out data dissemination. In other instances, private companies and academic institutions, as well as Federal, state, regional and local government agencies, can be additional conduits or points of sale for an agency's data. In some cases, this second group may serve as large a customer base as the agency's direct subscribers. Often resellers and redistributors help satisfy the agency's requirement to inform the public, while adding value to agency products. In many cases, these services cost the agency little or nothing.

Resellers and redistributors are key customers and their insights should be solicited and evaluated before an agency makes changes to its products.

Federal Resellers and Redistributors.--Several government organizations have established major centralized information redistribution services, to give the public technologically advanced, one-stop shopping opportunities for government information. Before embarking on the independent development of electronic information dissemination programs, statistical agencies should evaluate whether some of their programs could be carried out by a centralized service, thereby avoiding start-up costs and possibly reaching a wider audience.

- The National Technical Information Service (NTIS), U.S. Department of Commerce, is responsible for collecting, abstracting, indexing, archiving, reproducing, and disseminating scientific, technical, engineering, and business information. The NTIS FedWorld service can serve as a repository or a gateway for agency databases, software, and bulletin board systems. FedWorld is available via Internet or telephone modem. NTIS' FAX Direct service also offers fax-on-demand and broadcast fax services.
- STAT-USA, U.S. Department of Commerce, disseminates Federally-sponsored business and economic data. Services include the Economic Bulletin Board (EBB), an on-line service containing 3,400 files from all the major Federal economic agencies; EBB Fax, providing fax access to the most popular EBB files; the National Trade Data Bank CD-ROM, including 10,000 reports on international and export information; and the National Economic, Social, and Environmental Data Bank CD-ROM, containing information related to the U.S. economy, society, and environment.
- The U.S. Government Printing Office (GPO) has established an on-line electronic directory for Federal electronic information -- GPO Access; GPO is also developing model gateways for no-fee public access to databases in the GPO Access system via Internet, through selected depository libraries.
- The Depository Library System, itself, is the nation's oldest redistributor of government information, with government document repositories in more than 1,400 libraries in every state and territory.

Commercial Resellers and Redistributors.--Many private companies repackage, combine, or otherwise add value to an agency's products; then, they resell the information at a profit.

Private companies usually have a better feel for the marketplace than do governmental entities and are better able to tailor products for specific markets.

Agencies should consider the role played by private companies in the dissemination process. The Census Bureau, for example, maintains a list of private companies that provide services related to Census Bureau data and products. Companies "register" with the National Clearinghouse for Census Data Services and indicate the kinds of services they provide, the type and geographic extent of data they sell, and any specific or unique products they produce. The list is published as a brochure, reproduced in the Census Bureau Catalog, and otherwise made available to users who may require a different format or medium than that which the Census Bureau can provide. The Census Bureau does not franchise, inspect or endorse any one company over another.

Contracting-out Data Dissemination.-- Most agencies wish to maintain direct control over their data dissemination activities. However, several agencies have contracted out various pieces of the process.

- At the U.S. Geological Survey, general information requests and product sales are handled by an on-site contractor. Technical questions are referred to the appropriate USGS specialist.
- The Department of Agriculture's Economic Research Service and National Agricultural Statistics Service, as well as the Justice Department's Bureau of Justice Statistics, contract with a private company to sell products and handle subscription and mailing services.
- The Department of Housing and Urban Development, which provides all its products for free, contracts out the distribution process. Users can request free copies of publications, diskettes, and CD-ROMs from the service known as HUD User.

State and Local Redistributors.-- State and local agencies can also serve as distribution points for an agency's data. The extent to which nonfederal agencies take on this function varies from agency to agency.

Several agencies -- like the National Center for Education Statistics, the National Center for Health Statistics, and the Bureau of Labor Statistics -- have agreements with the states for joint data collection activities. In some cases, Federal monies fund state-level staff positions. These same state agencies also serve as local distribution points for the data. Other agency data products with state-level information are often provided as a matter of course.

The Census Bureau maintains cooperative agreements with each state and provides a designated state agency -- and through it, a network of affiliates -- any and all Census Bureau products for the state. Affiliates may be local planning agencies, councils of governments,

libraries, chambers of commerce, or university departments. This network, consisting of 1,800 organizations, responds to over one million information request a year. The program was recently expanded to include national minority organizations and their local affiliates.

Sources for Bibliographies and Abstracts

Although soon to be supplemented by the Government Information Locator System, centralized bibliographic and abstract services provide powerful supplementary locators to an agency's own information center. Agencies should verify that their products and services are adequately and accurately represented in these systems.

The following, as an example, are the principal sources of bibliographic and abstract references to energy statistical information:

- U.S. Government Printing Office -- Monthly Catalogue and Publications Reference File;
 - National Technical Information Service -- Bibliographic Database;
 - Office of Scientific and Technical Information -- Energy Science and Technology Database;
 - Congressional Information Service -- American Statistics Index and U.S. Government Periodicals Index; and
 - R. R. Bowker -- Energyline.
-

5: Cost and Financing of Electronic Dissemination

Federal agencies often find it difficult to assign a cost or price to their products and services -- especially when they are based on data collected for program administration or other agency purposes. This chapter provides some summary information on the budget and financing process and presents some guidelines on different options for costing electronic dissemination of statistical data.

When planning a new electronic dissemination effort, several key issues need to be part of the thinking in an agency's budget process. These include fixed costs -- such as software, hardware, and personnel overhead -- and variable costs -- like development, marketing, and customer service. The job is to estimate what these expenses will be and build them into the budget, so as to best provide new services to the broadest number of customers in the fairest manner.

General Cost Estimation

Much of the guidance on the classification of cost comes from the Office of Management and Budget. In recognizing that agencies have a responsibility to provide information to the public consistent with their missions, OMB Circular No. A-130 -- "Management of Federal Information Resources," says that an agency should consider the availability of information dissemination products from other sources equivalent to potential agency products, establish inventories and other finding aids, and provide notice and opportunities for consultation with respect to changes in information dissemination product offerings. In addition, agencies must take advantage of all dissemination channels, public and private, in discharging their dissemination responsibilities. The policies of the Circular were codified in the Paperwork Reduction Act of 1995, Public Law 104-13. The Circular also discourages exclusive arrangements made by or on behalf of an agency, including restrictions and imposition of fees on reuse or redissemination of Federal information.

In other words, the objective is to provide government information at the most reasonable cost (the lowest possible), while meeting the needs of users, based on the characteristics of the information. An agency should recover those costs that can be specifically identified as attributable to providing the dissemination service, but should not attempt to profit or recover other costs through the sale of Federal government information. A key factor in selecting the pricing approach is its relationship to the fulfillment of the agency's statutory mission.

Traditionally, government information providers have relied exclusively on printed forms of documents. Printed publications, press releases, and report summaries have been the principle products for most statistical organizations. When data users received these publica-

tions, they would often re-enter the information into some application that suited their needs. If the data users access the information directly in digital form, they can use the information more easily. They can also scan or search the information, and bring out relationships that would only be possible with considerable effort under the old system.

Today a considerable portion of the work prepared in a modern government statistical office originates in digital form. It would be safe to say that no press release, table, or publication exists solely in final typewritten form. At some time in its production cycle, the information passes through an electronic media for editing or processing. The key question then becomes how can this information be captured and made available to the public and to users in the most useful form? And, having done so, how can we recover expenses for that service? The answer to these questions lies in the classification of production costs. Careful consideration of the real boundary between work prepared for the agency mission and work performed for the dissemination activities will help define the appropriate role.

Fixed Costs

Certain costs of data dissemination exist regardless of the number of copies distributed. Data developers who package information must employ commonly used hardware and software, in order to be useful to the widest audience. The data provider should not require the data user to procure special software and hardware in order to use government information. Remember, while the set of software and hardware that is required by a "typical user" is constantly changing, the information provider should probably aim for a target slightly below the "cutting edge" of technology.

Software Development and Licensing Fees

If an agency wishes to provide information as something other than its original, raw form, it will be necessary to procure software needed to develop the data. This might take the form of special-purpose software -- such as CD-ROM authoring tools, communication packages, or compilers or other development tools needed to create the final product.

Costs for such products can vary widely, depending on the type of software used. For example, some Internet tools are publicly available at no cost. Other tools, such as development packages for particular interfaces, may require authoring software. Also, don't forget that costs of any individual package, itself -- in addition to any development tools -- must be included.

Some of the best statistical products will be useful to agency staff as well as the data user community. However, the agency may not always wish to provide access to its information in the same manner that its employees access the data. First, such information is probably in microdata form and is subject to privacy and confidentiality restrictions. Even if that is not the case, in-house access to information is often based on a mainframe orientation that was developed without any intention of being used by a less sophisticated audience. Most users of Federal statistics do not have the time, equipment, experience, or desire to study procedural

manuals in order to use the information properly in this format. They want information that has been customized for their needs.

Agencies may face difficulties in this area, since there is no unique solution that will fit all types of information. For example, a publication which is largely text-based would require a different sort of tool than one that is largely numeric. Finally, particularly in the area of software, agencies should be aware that, in the current environment, there is a good chance that whatever software is obtained may be obsolete by the time it is procured. That being the case, the best choice should be a good one which can be expected to have as long a shelf-life as possible.

Hardware Costs

In the past, hardware costs implied only mainframe expenses. Today, with the costs of most elements of electronic information dropping rapidly, hardware costs more typically relate to personal computers, mini-computers and related peripheral devices in connection with them. These costs have dropped dramatically in recent years and will probably continue to do so. In Chapter 2, many of these hardware requirements have been specified for the different media and cost estimates are provided.

Personnel Costs for Data Development

For information-based products, the cost of developing and collecting the data initially far outweighs the costs of distributing it. The single most important cost is time -- in terms of personnel used to put the information into usable forms. Even if the tools for packaging information are free, there will still be a substantial cost in collecting, assembling, organizing, maintaining, and updating the information.

Documentation of the Electronic Data Dissemination

Agencies should provide data users with enough information about the electronic dissemination system to access data. In the best circumstances, it will be better if they can provide the data in a manner that does not require users to purchase any special software. Integration across differing platforms should not be the data provider's problem. However, the data provider should be aware of significant differences across systems and attempt to provide the information in a form that does not prohibit any end user from making use of the information.

Remember, though, that documentation relates to the data and the tools or applications required to interface with them. If the agency develops its own package, it must also produce documentation on how the system works. (In some extreme cases, documentation on how to use a particular system can be larger than the data being accessed.) An alternative is to select standardized off-the-shelf products, which don't require extensive documentation. When doing so, however, care is always needed, since it can appear that the government is supporting one vendor over another if a particular platform is chosen.

Amortization of Fixed Costs

The amortization of fixed costs is a complex issue which may arise if an agency's goal is to have the revenue from the sale of information products cover dissemination costs (and

other costs). Typically, some set of fixed costs are identified (to be recovered) and then pricing is established to recover these costs by estimating demand (sales) over the life of the product. Alternatively, costs could be recovered in "standard accounting cycles" of five or ten years.

Controversy usually revolves around designating which costs should constitute fixed costs of the product and deciding whether they should be recovered. What part of the development costs are included -- hardware needed to produce the product, development software (e.g., authoring packages), labor for product development, marketing, and customer support? The inclusion of these items can greatly affect the per unit price -- especially with the small number of units that government information products usually disseminate. There are few guidelines about what costs should be included, and the practice varies widely from agency to agency. OMB Circular A-130 provides the most guidance in this area -- but it is also open to interpretation.

Variable Costs

Personnel

The costs of providing data to the public through electronic access fall most heavily in the area of personnel. Agency staff must be trained and knowledgeable about the information resources they are providing and be "the experts" for all questions relating to the information. Too often the coordination between the producers of the information and the branch releasing the data is minimal, with one side not fully aware of what the other is doing. Users calling in for help are unavoidably shuttled from one unit to another, while they experience a rising level of frustration and dissatisfaction with the information source. A fully implemented customer service program alleviates much of that problem, but adds substantially to personnel and training costs.

Customer Service

Two key areas of customer service cost deserve mention -- first is the need to provide broad technical support. As mentioned earlier, data users tend to have many questions. Obviously, there is a wide range of expertise with which the typical customer acquires and uses the information. Remember, however, that the data provider's responsibility is to the information provided, not to the platform on which the data are provided. Yet, inevitably, data producers find themselves answering question not about the data, but about the software used to acquire or to work with the data. While it is not the role of the government information services to provide advice about how to load files in *Lotus 1-2-3* or *Excel*, if the data are provided in that manner, the questions are unavoidable. In trying to be as responsive as possible to user needs, huge potential time costs may be incurred in translating from what users state they want to what they actually **need**.

Another area which can lead to escalating time costs is that of specialized requests or poorly defined inquiries. Many of the information requests that come into a Federal agency are very broad. Data users often request "everything you have on" without a clear or even

vague notion about how much is available and what would be involved in acquiring it. For some historical requests, there are largely mainframe-maintained databases. Luckily, many of these are being transferred to mini-computers and local area network environments, where they can be more readily accessed by the customer for analysis.

Billing Activities

If an agency elects to charge user fees for access to a product, then the responsibility for billing and handling of receipts must be assumed. Depending on the characteristics of the information product, this activity can produce a lot of overhead costs. For example, subscription products and on-line media incur more accounting overhead than products that are handled "on-demand." These billing activities require some degree of investment in computer systems, accounting procedures, and additional staff for accounting, communications (mail), and customer support.

Some agencies prefer not to collect user fees for products because of the expense of "handling money." It literally may be less expensive to give away the product than to sell it -- especially given the limited demand for many government information products. Other agencies elect to contract this function to outside organizations, such as the National Technical Information Service. Thus, the producing agency becomes a wholesaler, losing control of the dissemination of the product to some degree -- a less desirable situation, perhaps, since it may not be possible to recover the full cost of the product.

Free Products and Services

Several agencies have charged for diskettes, tapes, and large information requests. These charges were meant to discourage specialized, one-time requests, but were priced so that the cost of the recovery was covered for the request. Publicly-available software, such as *Mosaic* World Wide Web browsers, HTML editors, and postscript file viewers, are available for free, as are other tools for accessing information stored as *Adobe Acrobat* files, or Folio Views runtime readers. While 900-numbers exist to charge for electronically-accessed media, on-line media are harder to cost-out, so most agencies have opted not to charge for Internet access to data. On the other hand, there is no easy way to make frequently accessed materials available at zero or minimal costs. The user, of course, must pay to download files, but the agency cannot recover its expenses. Even data products which were prepared for internal or reproduction uses have additional costs associated with public dissemination -- the costs involved in maintaining a mailing list, mailing packages of information out, and so on.

A Caveat

In an electronic information environment, information that can be freely recopied will be recopied. Some of the leading data providers in the Federal government have had their information sources "mirrored" by other organizations. For example, the electronic bulletin board of the Department of Commerce is mirrored by the University of Michigan; the Bureau of Labor Statistics' system is mirrored by Sam Houston State University. In the past, the *Statistical Abstract of the United States* was a best-selling document. Since the information is government-funded, it cannot be copyrighted. After the government paid data development costs associated with the information, and packaged it into a usable format, private publishers

have taken the information, reproduced it exactly, and charged a price that was a fraction of the government's charge.

In the electronic information business, the replication of government data becomes even easier and the costs to the reproducer become even smaller. It is difficult to imagine how the government can recover its costs in this manner. As an example, the Department of Commerce's Economic Bulletin Board (EBB), established about 10 years ago, has the objective of recovering its costs through user fees. During this time, hardware and software costs have dropped dramatically. The system currently consists of a 486 personal computer with 48 modem lines, all of which costs about \$20,000 in today's market. The operation software costs about \$5,000. Peripherals for backup, etc., add another \$3,000 or so.

Since 1993 users have been able to access the EBB through Internet -- using the Telnet facility. Many users of this service "resell" information, typically after repackaging it. This activity has served to increase the utilization of the EBB. In recent years, however, large public institutions have begun "stripping" information from the EBB and duplicating the service verbatim. While wholesale replication of Federal information centers relieves some of the load placed on the agency's system, it also entails a loss of control and lack of accountability for errors and makes it impossible to recover data development costs.

Marketing and Promotion

A Federal statistical agency must also make its offerings known to the general public. Lacking resources to extensively promote its services, the typical Federal statistical agency must take advantage of low cost and/or nontraditional forms of advertising. Announcements of new products in agency publications and trade journals and presentations and demonstrations of them at academic conferences and conventions have been effective ways of promoting new products and services. Data on the Internet have been relatively easy to promote, due to the presence of lists and linkages that automatically route users to agency sources. Search software and browsers devoted to locating specific information also exist, so that users can ask general queries and be routed to statistical documents. Government-wide initiatives such as these should make it even easier for users to locate information.

To aid the public in locating and accessing government information, the Office of Management and Budget, in coordination with the Information Infrastructure Task Force, is promoting the establishment of an agency-based Government Information Locator Service. GILS will --

"identify information resources throughout the Executive Branch, describe the information available, and provide assistance in how to obtain the information. It will improve agencies abilities to carry out their records management responsibilities and to respond to Freedom of Information Act requests. It will also serve to reduce the information collection burden on the public by making existing information more readily available for sharing among agencies.

"GILS will consist of decentralized agency-based information locator records and associated information services. It will use off-the-shelf communications and information technology products and services, so that government information can be stored and retrieved in a variety of ways and in a variety of locations."

Agency responsibilities with respect to GILS are documented in OMB Bulletin 95-01, "Establishment of Government Information Locator Service." (This document is available on the Internet from <http://www.usgs.gov/public/gils/gilscopy.html>.)

Congestion Pricing

Electronic information access makes it possible to provide data in volumes that were previously impossible in paper publications. The detail that can be provided in a format such as a CD-ROM, for example, makes it possible to include large portions of material that would otherwise never see the public eye. If the information is bundled with other agency information and there's room available, the less popular information can piggy-back along with that which is in greater demand. Users with special needs for the obscure information get the best of all worlds -- information they've always wanted in digital form and other information the provider has added, as well.

On the other hand, as demand increases for on-line products, agencies will find that providing Internet access to their data is not free. As the size of information to be transferred becomes larger, the congestion costs of a single user can begin to burden the carrying capacity of the network. At present, larger data files can be released as physically transferable recorded media, instead, like diskettes or CD-ROM. Still, as new applications become more common, such as graphical data interfaces, multimedia packages, video, and audio, the size of the information vehicle becomes vastly larger.

Guidelines

With all of these different issues in mind, costing may seem more of a headache than it is worth. Following are some general guidelines to bear in mind when considering a new electronic dissemination system. (See also Figure E.) They should help budget and price products and services.

Agencies should establish an equitable approach to pricing, utilizing the simplest procedure possible. Subscription rates rather than actual usage are acceptable, as long as they can be demonstrated to be equally applied to all. Agencies may also want to consider amortizing overhead costs across all offered datasets, rather than establishing pricing for each based on usage.

When justified, the dissemination cost may include maintenance, training, customer support, computer and data communication cost for access to the data (or production cost for the delivered product, in the case of non-electronic techniques), and any additional cost associated specifically with providing the dissemination service. Some aspects of customer sup-

Figure E: Guidelines for Costing

- Build cost recovery into the budget process
- Develop a simple and fair pricing system
- Include all relevant costs in the price
- Define the collection process before the data are released
- Consider free or reduced rates, when reasonable
- Measure and analyze the results of cost recovery

port, for example, might be considered part of the mission responsibility and would not be included in the pricing mechanism. Developmental cost should not be included either.

Agencies should determine a method for collection of user fees and the frequency with which these fees will be collected before the product or service is made available. Particular attention must be made to handle non-payers and late payers.

Agencies may consider free or reduced rates to selected users. However, variable rates charged for different categories of users must be justified. Prices should be associated with the services provided, to the extent possible. Any intentional deviation should be open and explicitly explained to the public.

Agencies must have a mechanism in place to accurately measure access charges based on the established approach.

6: Latest Interagency Initiatives

The Subcommittee on Electronic Dissemination of Statistical Data held its first meeting in January 1994. At that time, the first graphical browser for the World Wide Web (WWW) -- *Mosaic* -- was known only to a handful of people involved in its early testing stages. As this report goes to print, we find that the major statistical agencies' current dissemination priorities are through the WWW. The WWW has already increased the number of users of government statistics many-fold and the growth is likely to continue. We hope that statistical agencies not yet disseminating information through the WWW will find this report helpful in their transition to greater electronic dissemination.

Several new interagency initiatives are underway, which will likely determine the future development and direction for cooperative work among statistical agencies interested in improving the electronic dissemination of statistical data. These initiatives include --

- **Interagency Internet Group.**-- Representatives from several statistical agencies meet quarterly to exchange ideas, discuss new initiatives, and determine areas where inter-agency cooperation would be beneficial.
 - **Census- and BLS-sponsored Metadata Workshop, November 1995.**-- As discussed in this report, dissemination of metadata -- or data about the data -- is an important consideration in planning for electronic dissemination. This workshop included presentations on standards for statistical and geo-spatial metadata, guidelines on the "best" ways to access information, and discussions on the relationships between on-line metadata and printed information.
 - **Economic and Social Indicators Briefing Rooms.**-- Recently the President's Office of Science and Technology Policy, along with representatives of the major statistical agencies, began developing an Economic Briefing Room, which will be accessible through the White House WWW home page. This initiative will provide the general public easy access to key current economic indicators, with links to the agencies that produced these indicators. Plans for developing a Social Indicators Briefing Room are also underway.
 - **Interagency Council on Statistical Policy's Task Force on One-Stop Shopping for Statistical Data.**-- The One-Stop Shopping Task Force was established by the Office of Management and Budget's Interagency Council on Statistical Policy. The Task Force is primarily responsible for determining an overall approach to the design and development of easy WWW access to all Federal statistics. The Task Force is expected to obtain input from important stakeholders, including representatives from the Association of Public Data Users (APDU), the Council of Professional Associations on Federal Statistics (COPAFS), academic researchers, and statistical agencies
-

not represented on the Task Force. The Task Force is also expected to take advantage of recent scientific research activities related to its mission, especially digital library research.

Members of the Subcommittee on Electronic Dissemination of Statistical Data have been and will continue to be active in forums that will help increase the usefulness of this report. Members have already participated in sessions at meetings of the American Statistical Association and APDU. There are also plans to work with COPAFS and to speak at their May 1996 conference on the quality of Federal statistics. While the Subcommittee's formal work is over, the members share a commitment to increased electronic access to Federal statistical data and welcome suggestions for other ways to help in improving the dissemination of Federal statistics.

References

- Energy Information Administration (1995). *Energy Information Directory*, Department of Energy, Washington, DC (DOE/EIA/0205).
- Office of Management and Budget (1995). *Establishment of Government Information Locator Service*, OMB Bulletin 95-01.
- Office of Management and Budget (1994). Notice: Management of Federal Information Resources, OMB Circular No. A-130 in *Federal Register*, 59, 141, 37906-37928.
- Office of Management and Budget (1994). *Schedule of Release Dates for Principal Federal Economic Indicators for 1995*.
- Office of Management and Budget (1993). *Resource Manual for Customer Surveys*, Statistical Policy Office, Office of Information and Regulatory Affairs.
- Paperwork Reduction Act of 1995*, Public Law 104-13.
- President of the United States (1993). Executive Order 12862 on Setting Customer Standards, September 11, 1993.
- Subcommittee on Statistical Disclosure Limitation Methodology (1994). *Report on Statistical Disclosure Limitation Methodology*, Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, Statistical Policy Working Paper 22.
- Subcommittee on Statistical Disclosure and Disclosure Avoidance (1978). *Report on Statistical Disclosure and Disclosure Avoidance Techniques*, Federal Committee on Statistical Methodology, Office of Federal Statistical Policy and Standards, U. S. Department of Commerce, Statistical Policy Working Paper 2.
- U.S. Bureau of the Census (1994). *Statistical Abstract of the United States: 1994*, 114th edition, Washington, DC.
-

Appendix

This Appendix is intended to describe the current status, as of October 1995, of electronic dissemination of information by Federal statistical agencies. Readers of this report in late 1995 and early 1996 may find it useful to benchmark their practices against the practices described in the Appendix. To continue to be useful, the Appendix will need to be updated periodically. Please check the WWW site for this report at the National Science Foundation (<http://www.nsf.gov/sbe/srs/stats.htm>) for updates.

In order to obtain the information for this section, about 20 Federal statistical agencies were contacted. They were asked to describe their current status (as of October 1995) with regard to electronic dissemination of statistical data and to provide contact information for the various media they produce. They were also asked to fill in a grid, indicating the kinds of media available from their agency. The resulting information has been summarized in three different ways --

- Figure A1 lists Internet addresses for 21 major statistical agencies. Note that all agencies have a WWW site, but not all offer FTP and Gopher service.
- Figure A2 is a grid, which summarizes the types of media each of the responding agencies offer at this time.
- Agency Contacts and Summary of Current Practices -- The balance of the Appendix provides more specific information -- by agency -- regarding their dissemination of statistical data. Details about the kinds of data available and contact numbers to obtain information about these data are provided.

In addition to serving as a benchmark, this Appendix should also be viewed as a resource for customer service centers for reference purposes.

Figure A1.--Internet Addresses for Federal Statistical Agencies

<p>Agency for Health Care Policy and Research FTP: //ftp.ahcpr.gov Gopher: WWW http: //www.ahcpr.gov</p> <p>Bureau of the Census FTP: //ftp.census.gov/pub Gopher: //gopher.census.gov:70 WWW http: //www.census.gov</p> <p>Bureau of Economic Analysis FTP: Gopher: WWW http: //www.doc.gov/resources/beahome.html</p> <p>Bureau of Justice Statistics FTP: Gopher: //justice2.usdoj.gov:70/11/ojp/bjs WWW http: //ncjrs.aspensys.com:81/aboutbjs.html</p> <p>Bureau of Labor Statistics FTP: //stats.bls.gov Gopher: //stats.bls.gov WWW http: //stats.bls.gov</p> <p>Bureau of Transportation Statistics FTP: //ftp.bts.gov Gopher: //gopher.bts.gov WWW http: //www.bts.gov</p> <p>Centers for Disease Control & Prevention FTP: //ftp.cdc.gov Gopher: //gopher.cdc.gov WWW http: //www.cdc.gov</p> <p>Economic Research Service, Dept. of Agriculture FTP: //usda.mannlib.cornell.edu Gopher: //usda.mannlib.cornell.edu WWW http: //www.econ.ag.gov</p> <p>Energy Information Administration FTP: //ftp.eia.doe.gov Gopher: //gopher.eia.doe.gov WWW http: //www.eia.doe.gov</p> <p>Environmental Protection Agency FTP: //ftp.epa.gov Gopher: //gopher.epa.gov WWW http: //www.epa.gov</p> <p>Div. of Research & Statistics, FDIC FTP: //ftp.fdic.gov Gopher: //gopher.fdic.gov WWW http: //www.fdic.gov/fdicdrs.html</p>	<p>FedWorld at National Technical Information Service FTP: Gopher: WWW http: //www.fedworld.gov</p> <p>Health Care Financing Administration, SSA FTP: Gopher: WWW http: //www.ssa.gov/hcfa/hcfahp2.html</p> <p>National Agricultural Statistics Service FTP: //usda.mannlib.cornell.edu Gopher: //usda.mannlib.cornell.edu WWW http: //www.usda.gov/nass</p> <p>National Center for Education Statistics FTP: //ftp.ed.gov Gopher: //gopher.ed.gov WWW http: //www.ed.gov/pubs/stats.html</p> <p>National Center for Health Statistics FTP: Gopher: WWW http: //www.cdc.gov/nchswww/nchshome.htm</p> <p>Office of Research & Statistics, SSA FTP: //ftp.ssa.gov/pub Gopher: //gopher.ssa.gov WWW http : //www.ssa.gov/statistics/ors_home.html</p> <p>Science Resources Studies, National Science Foundation FTP: //stis.nsf.gov Gopher: //stis.nsf.gov WWW http: //www.nsf.gov/sbe/srs/stats.htm</p> <p>Statistics of Income, Internal Revenue Service FTP: Gopher: WWW http: //www.ustreas.gov/treasury/bureaus/irs/irs.html</p> <p>STAT-USA, Commerce FTP: Gopher: WWW http: //www.stat-usa.gov</p> <p>U. S. Postal Service FTP: Gopher: WWW http: //www.usps.gov</p>
Adapted from "Statistical Agencies on the Internet" (10/20/95) by Dan Rope (Rope_D@bls.gov), Bureau of Labor Statistics.	

Figure A2.--Summary of Media Availability by Agency

Agency	Electronic Bulletin Boards	Faxes & Auto Faxback	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
Agency for HealthCare Policy & Research	X	X	X	X	X	X	X	
Bureau of the Census	X	X		X	X	X	X	X
Bureau of Economic Analysis	X	X		X	X	X	X	X
Bureau of Justice Stat.	X	X		Planned	X	X	X	
Bureau of Labor Stat.	X	X	X	X	X	X	X	
Bureau of Transportation Stat.	X	X		X	X	X		
Centers for Disease Control & Prevention	X	X	X	X	X	X	X	X
Economic Research Service, Dept. of Agriculture	X	X		X	X	X		
Energy Information Administration	X	X	Planned	X	X	X	X	
Federal Reserve Board	X	X			X	X	X	
Immigration and Naturalization Serv.		X		X		X	X	
National Agricultural Statistics Service	X	X		X	X	X		X
National Center for Education Statistics	X		X	X	X	Limited	X	X
National Center for Health Statistics			X	X	X	X	X	X
Office of Research & Statistics, SSA		X		X		X	X	
Science Resources Studies, Nat'l Science Foundation	X	X	X	X	X	X	X	
Statistics of Income, IRS	X	X		Planned		X	X	X
STAT-USA, Commerce	X	X		X	X			
U. S. Geological Survey	X	X	X	X	X	X	X	X
U. S. Postal Service				X		X		

**Agency for Health Care Policy and Research
(HHS/AHCPR)**

Agency for Health Care Policy and Research
Executive Office Center
2101 E. Jefferson St., Ste. 501
Rockville, MD 20852

General Contact Information:

Deirdre Herman, Phone: (301) 594-1364 x 165, e-mail: dherman@po5.ahcpr.gov; for Healthcare Cost and Utilization Project, hcupnis@cghsir.ahcpr.gov, or contact Roxanne Andrews, randrews@cghsir.ahcpr.gov, (301) 594-1410

In August 1995, AHCPR released the Healthcare Cost and Utilization Project (HCUP-3) Nationwide Inpatient Sample (NIS), Release 1, which contains data on all inpatient hospital stays from a 20% sample of U.S. hospitals. HCUP-3 is a Federal-state-industry partnership to build a multi-state health care database. HCUP-3 was the first microdata product made available by AHCPR in both magnetic tape and CD-ROM format. The NIS, Release 1 includes data from 1988-1992. NIS, Release 2, which will include data from 1993, is expected to be released in early 1996.

Other Agency products include data from the AIDS Cost and Services Utilization Survey (ACSUS) and the 1987 National Medical Expenditure Survey (NMES). ACSUS is a longitudinal study of persons with HIV-related disease and is one of the largest data collection efforts targeting the population of persons infected with HIV. NMES continues a series of national health care expenditure surveys, most recently the 1980 National Medical Care Utilization and Expenditure Survey and the 1977 National Medical Care Expenditure Survey. NMES Household Survey covers both the U.S. civilian, noninstitutionalized population, as well as residents and admissions to nursing homes and facilities for the mentally retarded, over the course of the survey year.

Both HCUP-3 and NMES are sources for numerous AHCPR publications of statistical data in tabular form. These statistical reports cover a broad range of topics, including national estimates of the cost of inpatient hospital care, insurance coverage of the population, the number of uninsured in the population, use of health services by the insured and uninsured, as well as national estimates of total health expenditures.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	YES	YES	YES	YES	YES	NO
Sale Outside	Free In-house	Sale Outside	Free In-house	Sale Outside	Sale Outside	Sale Outside	
dherman@po5.ahcpr.gov (301) 594-1364 x 165	smerrill@po5.ahcpr.gov (301) 594-1364 x 149	NTIS FED WORLD dherman@po5.ahcpr.gov	Same	Same	Same	Same	
Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.							

**Bureau of the Census
(DOC/CENSUS)**

Bureau of the Census
Customer Services
Room 3649, FB 3
Washington, DC 20233

General Contact Information:

Customer Services, Washington, DC 20233, phone: (301) 457-4100, fax: (301) 457-4714,
e-mail: **gatekeeper@census.gov**

General subject matter categories of data offering at the Census Bureau include: demographic data, economic data, small area data, product information, press releases and research papers, geographic data, and custom data extract capabilities. Some specific examples of data include: population estimates and projections; economic indicators; international trade data; state rankings and profiles, and financial data for states, counties, cities, and school districts.

Most of the data are available on the Internet and optical disc (CD-ROM). Some of the data are available through electronic bulletin board, faxes and auto faxbacks, data tapes, and cartridges. Diskettes are rarely used to disseminate electronic information. Information on the electronic bulletin board, faxes and auto faxbacks are provided in text and/or tabular format. All other media provide data in text, tabular and/or microdata formats.

As part of the Bureau's effort to dramatically expand Americans' access to official demographic and economic information, the Census Bureau plans to expand electronic dissemination of its data. The Internet, along with other electronic systems, will gradually become the primary source for Census Bureau statistics.

The Bureau is planning to design, develop, and implement a system for data access and dissemination called DADS. The objective is to provide one general (electronic) system for all access to Census Bureau data. This system will be fast, flexible, and cost-efficient. The initial guiding principles are that geography is the integrating factor, Title 13 data are protected, metadata will be available integrally with the data, and the number of pre-specified printed reports will be minimized.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	YES	YES	YES	YES	YES
Free In-house	Free/Sale In-house		Free In-house	Sale In-house	Sale In-house	Sale In-house	Sale In-house
Customer Services (301) 457-4100	Same		Same	Same	Same	Same	Cartridges Same

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

**Bureau of Economic Analysis
(DOC/BEA)**

Bureau of Economic Analysis
1441 L. St. NW
Room 6044, Stop BE-62
Washington, DC 20230

General Contact Information:

BEA REIS CD-ROM: (202) 606-5360; Survey STAT-USA CD-ROM: (202) 606-9695; e-mail: leland.scott@bea.doc.gov

Most of BEA's economic statistics are made available in electronic form, with diskettes being the most widespread medium used. In addition to diskettes, on-line access -- both by bulletin board and by Internet (World Wide Web) -- is provided for BEA's news releases and for estimates from BEA's national income and product accounts, its "leading indicators" program (which will be transferred to the Conference Board later in 1995), and its regional economic accounts. All of these statistics are made available as ASCII tables; most of the diskettes also include a program that allows users to import the ASCII tables into *Lotus 1-2-3* spreadsheets. BEA's monthly journal, *Survey of Current Business*, is published on the Internet and on CD-ROM in *Adobe's* portable document format; the CD-ROM includes a utility that allows *Windows* users to copy tables into spreadsheet format. A few BEA databases are still made available on magnetic tape, but BEA is considering discontinuation of this medium and use of an alternative, perhaps write-once CD-ROMs. Most of the detailed estimates from BEA's regional accounts are published on an annual CD-ROM. In addition to on-line access, BEA's news releases are available by fax, and summaries are available as recorded telephone messages. BEA also publishes most of its major databases on two CD-ROMs produced by STAT-USA; these databases are in the National Trade Data Bank in ASCII format.

In the coming year, BEA expects to broaden the availability of its estimates on the Internet; as part of this effort, BEA will experiment with designing an interactive system of user-access to its data. BEA also anticipates somewhat greater use of CD-ROMs, particularly as an alternative way to disseminate print publications.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	YES	YES	YES	YES	YES
Sale Outside	Sale Outside		Sale Outside	Sale In/Outside	Sale In-house	Sale In-house	Free In-house
STAT-USA EBB	STAT-USA FAX		STAT-USA WWW leland.scott@bea.doc.gov (202) 606-9695	NTIS, GPO	USER'S GUIDE TO BEA INFORMATION	USER'S GUIDE TO BEA INFORMATION	Recorded Telephone Messages (202) 606-9900

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

Bureau of Justice Statistics
633 Indiana Ave., NW
Washington, DC 20531

Electronic data -- Crime and Justice Programs, Inter-university Consortium for Political and Social Research, P.O. Box 1248, Ann Arbor, MI 48106-1248, phone: (313) 763-5011, e-mail: cdunn@icpsr.umich.edu; Publications -- BJS Clearinghouse, P.O. Box 6000, Rockville, MD 20850, phone: (301) 251-5242

All BJS data files are available on tape from the National Archive of Criminal Justice Data, operated by the Inter-university Consortium for Political and Social Research. Smaller files are also available on diskette, and topical CD-ROMs containing multiple data files are also prepared. Files are available free to ICPSR-member institutions, analysts in state and local government, and researchers working on DOJ research projects; they are available to all others on a cost-reimbursable basis from ICPSR. CD-ROMs are available to non-members from the National Criminal Justice Reference Service for \$25.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	Planned	YES	YES	YES	NO
Free Outside	Free Outside		1996	Free/Sale Outside	Free/Sale Outside	Free/Sale Outside	
NCJRS K. Crisafulli (301) 251-5242	NCJRS (301) 216-1827		NACJD cdunn@icpsr.u mich.edu (301) 764-5011	NCJRS NACJD K. Crisafulli C. Dunn	NACJD C. Dunn	NACJD C. Dunn	

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

Bureau of Labor Statistics
(DOL/BLS)

Bureau of Labor Statistics
2 Massachusetts Avenue, NE
Room 4110
Washington DC 20212-0001

General Contact Information:

Customer Service, 2 Massachusetts Ave. NE, Room 2850, Washington DC 20212-0001, phone: (202) 606-5886, e-mail: **McMichael_M@bls.gov**

The Bureau of Labor Statistics provides current and historical data, principally in the area of labor economics, in electronic form. Among the categories of data available to the public are employment data from the household and payroll surveys, the Producer Price Index, the Consumer Price Index, the Employment Cost Index, productivity data, statistics on the provision of employee benefits, and data on the incidence of occupational injuries and illness. These statistics are available through LABSTAT, the Bureau's public database, that contains both current and historical data, as well as current press releases.

The Bureau's present focus is on providing all the data possible through the Internet, and more specifically, the World Wide Web. Our immediate goal was to put up the most frequently requested information and documents, and this was essentially completed on Labor Day 1995. Our short-term goals are two-fold: to convert our production process, so that we are producing electronic layout as well as paper layout; and to assist the user in converting our raw data to information, through the use of custom charts, graphs, and other analytical tools. Our long-term goal will be to capture 111 years of published material and make it available to the public in the most effective means possible.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	YES	YES	YES	YES	YES	NO
Free In-house	Free In-house	Free In-house	Free In-house	Sale In-house	Sale In-house	Sale In-house	
McMichael_M@bls.gov	Same	Same	Same	Same	Same	Same	

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

Bureau of Transportation Statistics
(DOT/BTS)

Bureau of Transportation Statistics
Room 3430, 400 Seventh Street, SW
Washington, DC 20590

General Contact Information:

Phone: (202) 366-3081; fax: (202) 366-3640; e-mail: **robert.zarnetske@bts.gov**; Customer Services: Att. Kathleen Bradley, Room 3430, 400 Seventh St, SW, Washington, DC 20590, phone: (202) 366-3282, fax: (202) 366-3640, e-mail: **kathleen.bradley@bts.gov**

The Bureau provides information and data that describe the state of the nation's transportation system. The Bureau provides tabular data in ASCII and various spreadsheet formats, microdata in database formats, and various textual reports. Every product produced by the Bureau is available in electronic form, on the Internet, CD-ROM, or diskette.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	YES	YES	YES	NO	NO
Free In-house	Free In-house		Free In-house		Free In-house		
(800) 363-4BTS	(800) 671-8012		(Full access) www.bts.gov	Includes 22 titles; multi-media			

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

Economic Research Service
1301 New York Avenue, NW
Washington, D.C. 20005-4788

ERS Information Center, 1301 New York Ave., NW, Room 110, Washington, DC 20005-4788, phone: (202) 219-0515, fax: (202) 219-0112, e-mail: service@econ.ag.gov

ERS continues to expand the quantity and scope of agency data available electronically in all media. As the capabilities of the Internet develop, the agency expects to offer interactive databases on the Internet that may not be offered through other media.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	YES	YES	YES	NO	NO
Free In-house	Free In-house		Free In-house	Sale In-house	Sale In-house		
ERS/NASS BBS (800) 821-6229 9600 BAUD, (N, 8, 1) pkorb@econ. ag.gov (202) 219-0758	AutoFax (202) 219-1107 Information Center service@ econ.ag.gov (202) 219-0515		Gopher usda.mannlib. cornell.edu WWW www.econ.ag. gov	ERS/NASS Stock # 95050 (800) 999-6779	Same		

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

**Energy Information Administration
(DOE/EIA)**

Energy Information Administration
U.S. Department of Energy
Washington, DC 20585

General Contact Information:
National Energy Information Center, phone: (202) 586-8800, e-mail: **infoctr@eia.doe.gov**

The Energy Information Administration (EIA), the world-wide leader in energy data, analysis, and forecasting, makes available in various electronic formats statistical data and information on energy resources and reserves, energy production, energy demand and consumption, energy technologies, and related economic and statistical information relevant to the adequacy of energy resources to meet the Nation's energy demands in the near- and longer-term future.

EIA's electronic information takes the form of descriptive and analytic text; tables of current and historic time series and forecast data; public-use data files and databases; and combined database/software applications that permit the user to generate results. Dissemination media include facsimiles, CD-ROMs, Internet servers, bulletin boards, diskettes, and tapes.

EIA's future plans include increasing the availability of information on the Internet; publishing a periodical CD-ROM containing the bulk of EIA's publications, databases, and applications; instituting a fax-on-demand service in 1996; and creating a corporate electronic file repository accessible to both Internet and dial-up customers.

EIA currently disseminates 5 reports via broadcast fax: 4 seasonal "watches" (Low Sulfur Distillate Watch, RFG Watch, Propane Watch, and Petroleum Market Report) and the Weekly Coal Report.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	PLANNED 1996	YES	YES	YES	YES	NO
Free/Sale In/Outside	Free/Sale In/Outside		Free In-house	Free/Sale In/Outside	Sale Outside	Sale Outside	
NEIC infoctr@ eia.doe.gov (202) 586-8800	Same		Same www.eia.doe.gov	Same	Same	Same	

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

Federal Reserve Board
20th & Constitution Avenue, NW
Washington, DC 20551

Publication Services, Federal Reserve Board, Washington, DC 20551, phone: (202) 452-3245, fax: (202) 728-5886

Data available in electronic form from the Federal Reserve fall into the following categories:

- Banking statistics
- Statistics on the general economy and financial markets
- Statistics on interest rates and lending
- Tapes of commercial bank call and income reports (DOC-NTIS)
- Home Loan Disclosure Act data on mortgage lending (CD-ROM: Federal Financial Institutions Examination Council)
- Texts of testimony before Congressional committees, minutes of meetings of the Federal Open Market Committee, and District Economic Conditions, (*The Beige Book*) (DOC Bulletin Board)

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	NO	YES	YES	YES	NO
Sale Outside	Free/Sale In-house			Sale Outside	Sale In-house	Sale Outside	
NTIS DOC-EBB				FFIEC		NTIS	

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

**Immigration and Naturalization Service
(DOJ/INS)**

Immigration and Naturalization Service
Statistics Division
425 Eye St, NW
Washington, DC 20536
Attn: Tariff Bldg Room 268

General Contact Information:

Phone: (202) 376-3066, fax: (202) 376-3083, e-mail: hoefermi@justice.usdoj.gov

The INS provides a limited number of data series through electronic dissemination. Data on legal immigration is provided on tape to interested users through the National Technical Information Service (NTIS). We also have tapes relating to legalized aliens available at NTIS. Other data series that are in less demand are sold on tape or sometimes as PC microdata files by the Statistics Division.

INS intends to make more information available through the Internet in text and tabular form. A report on legal immigration with tables was posted recently on the Department of Justice Gopher.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
NO	YES	NO	YES	NO	YES	YES	NO
	Free In-house		Sale In-house		Free/Sale In-house	Sale In/Out	
	NTIS		NTIS		NTIS	NTIS	

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

National Agricultural Statistics Service
Room 5805 - South Agriculture
14th & Independence Ave., SW
Washington, DC 20250-2000

Phone: (800) 727-9540 or (202) 720-7017, fax: (202) 690-1311, e-mail: NASS@AG.GOV

All scheduled statistical reports are loaded to the USDA Computerized Information Delivery Service (CIDS). CIDS operates as a commercial bulletin board with monthly fee and usage charges. A wide variety of reports from other USDA agencies, including a number of reports dealing with foreign agricultural production and export opportunities are also available over CIDS. Selected NASS reports are loaded to a free ERS-NASS bulletin board. A limited amount of fax transmissions are made from NASS Headquarters; however some NASS state statistical offices do use autofax for time-sensitive reports to Headquarters.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	YES	YES	YES	NO	YES
Free/Sale In/Out	Free/Sale In-house		Free In-house	Sale In/Out	Sale In-house		Free In-house
ERS/NASS EBB	ERS/NASS Fax		Cornell University Mann Library	<i>Agricultural Statistics</i>	NASS@AG.GOV		News Media Lock-up Procedures P. Joyce Hotline: (800) 727-9540 (202) 720-7017

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

**National Center for Education Statistics
(Ed/NCES)**

National Center for Education Statistics
555 New Jersey Ave., NW
Washington, DC 20208-5652

General Contact Information:

National Data Resource Center (NDRC) 1900 N. Beauregard St., Suite 200, Alexandria, VA 22311-1722, phone: (703) 845-3151, fax: (703) 820-7465, e-mail: ndrc@inet.ed.gov

The National Center for Education Statistics (NCES) collects statistics on the condition of education in the United States, analyzes and reports the meaning and significance of these statistics, and assists states and local education agencies in improving their statistical systems. NCES makes its data available in a variety of electronic formats, primarily using CD-ROM and the Internet. NCES provides major databases, electronic codebooks and table generation software on CD-ROM for general use, and routinely releases data via CD-ROM through the Government Printing Office. NCES uses the Internet for dissemination of files and publications, including announcements of new publications and data sets, descriptions of programs, press releases, searchable directories, funding opportunities, event calendars, and full-text publications.

NCES provides its customers with the ability to design and request their own tables and correlation matrices through the Internet using its Data Analysis System (DAS) server. NCES has established the National Data Resource Center (NDRC) to broaden access to data and information from studies and surveys that NCES maintains. In addition to NCES projects which are available off-the-shelf, the NDRC provides various customer services to enhance the availability of NCES data, including data in customized formats and/or on customized media.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	NO	YES	YES	YES	YES	YES	YES
Free In-house		Free In-house	Free In-house	Sale Outside	(Limited)	Sale In-house	Free In-house
EBB (800) 222-4922 max 14,400 baud		EBB (202) 219-1511 max 14,400 baud	www.ed.gov	Superintendent of Documents (202) 512-1800		NLERSCD (202) 219-1522	NDRC ndrc@inet.ed.gov (703) 845-3151

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

National Center for Health Statistics
6525 Belcrest Rd., Rm. 1050
Hyattsville, MD 20872

Data Dissemination Branch, Division of Data Services, National Center for Health Statistics, 6525 Belcrest Rd., Room 1064, Hyattsville MD 20782, phone: (301) 436-8500, fax: (301) 436-4258, e-mail: **nchsquery@nch10a.em.cdc.gov**

In addition to a wide range of printed materials, NCHS releases information through the following means: floppy diskettes contain survey data as well as tabular and textual materials published in NCHS reports; public-use data tapes contain survey microdata as well as survey documentation; CD-ROMs contain all information on the public-use data tapes as well as NCHS' own software, provided to retrieve, access and search the data and documentation; the NCHS home page on the Internet World Wide Web contains information about NCHS programs and services, press releases and fact sheets, viewable and downloadable files of NCHS publications, and an interactive query function. Furthermore, NCHS releases its data on-line through WONDER and PC-WONDER, which is developed and maintained by the Centers for Disease Control and Prevention in Atlanta. NCHS has developed the Statistical Export and Tabulation System (SETS), which provides developers with tools to produce CD-ROMs or diskettes containing databases and documentation for distribution. The SETS Designer Kit is being used by private and public organizations and is for sale by GPO.

Plans include print-on-demand services, Internet links to state health data, and improved access.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
NO	NO	YES	YES	YES	YES	YES	YES
		Free In-house	Free In-house	Sale Outside	Sale Outside	Sale Outside	Free/Sale In/Out
		CDC WONDER (770) 488-7536	Data Services www.cdc.gov/ nchswww/ nchshome.htm	NTIS GPO	NTIS GPO	NTIS GPO	SETS Software NTIS GPO

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

National Science Foundation
Division of Science Resources Studies
4201 Wilson Boulevard, Suite 965
Arlington, VA 22230

Division of Science Resources Studies -- phone: (703) 306-1780, fax: (703) 306-0510; Publications -- e-mail: pubs@nsf.gov; NSF's Science and Technology Information System (STIS) -- phone: (703) 306-0212; CASPAR -- Quantum Research Corporation, 7315 Wisconsin Avenue, Suite 631 W, Bethesda, MD 20814, phone: (301) 657-3070

In addition to producing print publications, SRS publications and tables are made available through the SRS home page on the World Wide Web (<http://www.nsf.gov/sbe/srs/stats.htm>) and through the National Science Foundation's Science and Technology Information System (STIS), which is accessible through anonymous ftp, telnet, gopher, e-mail, and non-Internet remote log-in via modem. STIS features a full-featured text search and retrieval software (TOPIC) to help users locate documents. SRS also maintains CASPAR, a database system on compact CD-ROM, containing information about academic science and engineering resources. CASPAR files are also available over the Internet.

Current plans for enhancements to its electronic dissemination include revisions to CASPAR, to enhance its user friendliness, and development of a user-friendly database system in which individuals can produce custom-designed tabulations from SRS work force data. This latter system is known as Scientist and Engineer Statistics Data System (SESTAT). SRS is also working on increasing the amount of metadata available related to its surveys that are accessible on the Internet. With the increasing popularity of the World Wide Web, SRS has been devoting more resources to preparation of electronic documents in *hypertext* form for posting on the Web.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	YES	YES	YES	YES	YES	NO
Free In-house	Sale In-house	Free In-house	Free In-house	Sale In-house	Free In-house	Free/Sale In/Outside	
STIS (703) 306-0212 1200,2400,9600 baud 7-E-1 Login: Public	Fax (703) 644-4278	STIS (703) 306-0212 1200,2400,9600 baud 7-E-1 Login: Public	www.nsf.gov/ sbe/srs/stats.htm	CASPAR (301) 657-3070	(703) 306-1780	Same	

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

Statistics of Income
Internal Revenue Service
(IRS/SOI)

Statistics of Income Division (CP:R:S:P)
Internal Revenue Service
P.O. Box 2608
Washington, DC 20013-2608

General Contact Information:

Statistical Information Services, Statistics of Income (CP:R:S:P), PO Box 2608,
Washington, DC 20013-2608, phone: (202) 874-0410, fax: (202) 874-0922, e-mail: **beth.kilss@wpgate.irs.gov**

Statistics of Income (SOI) of the Internal Revenue Service produces detailed income and tax statistics for use by tax researchers, demographers, and economic analysts. SOI is primarily responsible for obtaining, tabulating, and publishing tax return statistics and for providing the Department of the Treasury and Congress with information on income and taxes for use in analyzing existing and proposed tax policies. The data also are used by the Commerce Department's Bureau of Economic Analysis in the development of the national accounts. Tax information available include data for individuals, corporations, partnerships, sole proprietorships, tax-exempt organizations, estate tax and personal wealth studies, foreign activity of U.S. corporations and citizens, and activity of foreign corporations in the U.S.

All SOI reports are available in hardcopy. Selected data from them -- including text and tabular material from recent issues of the quarterly *SOI Bulletin* -- are among the 1,000 or so files now available on the SOI Electronic Bulletin Board (EBB). Faxes and auto-faxbacks of publicly-available information can be requested on a limited basis. CD-ROMs of historical data on individuals are being prepared under contract and will be for sale in the future. Most public-use microdata and tabular data are available on a reimbursable basis on diskette, magnetic tape, or microfilm. Plans are underway to establish an Internet node for SOI, which would give users greater manipulability with regard to SOI data. The SOI EBB is currently accessible through FedWorld, IRS-IS BBS, IRS Forms BBS, and soon the IRS World Wide Web home page and the Treasury World Wide Web home page.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	Planned for 1996	NO	YES	YES	YES
Free In-house	Free/Sale In-house				Sale In-house	Sale In-house	Sale In-house
Sign-on: (202) 874-9574	Statistical Information Services Staff (202) 874-0410				Same	Same	Micro-film
Administrator: (202) 874-0277							Same

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

STAT-USA
Commerce
(DOC/STAT-USA)

STAT-USA
H4885
U.S. Department of Commerce
Washington, D.C. 20230

General Contact Information:
Phone: (202) 482-1986; fax: (202) 482-2164; e-mail: stat-usa@doc.gov

STAT-USA is the official source for electronic releases from the Economics and Statistics Administration (ESA) within the U.S. Department of Commerce. The STAT-USA "repository" includes data and press releases from ESA on a number of important topics, such as national economic indicators and related analytical reports from the Chief Economist of the United States. In addition, STAT-USA provides a "Federal government-wide" one-stop "shop" for data and information of interest to business and economics users. The official home of the National Trade Data Bank (a valuable information source for businesses involved in foreign trade) is STAT-USA, but over forty agencies contribute to this database on a regular basis. In addition, many agencies use STAT-USA as their electronic dissemination "partner" in providing information to the public. Such diverse agencies as the Federal Reserve Board, the National Labor Relations Board, and the Agency for International Development release their electronic information through STAT-USA.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	NO	YES	YES	NO	NO	NO
Sale In-house	Sales In-house		Sales In-house	Sales In-Outside			
F. Williams (202) 482-3429	Same		P. Christy (202) 482-0123	F. Williams (202) 482-3429			
fwilliams@doc.gov			pchristy@doc.gov	fwilliams@doc.			

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

**U.S. Geological Survey
(USGS)**

U.S. Geological Survey
508 National Center
Reston, VA 22092

General Contact Information:
Phone: (800) 872-6277, fax (703) 648-5548

The U.S. Geological Survey distributes statistics data in the areas of geology (hazards, resources, and environment), water (surface water quantity and flow, surface and ground water quality), and basic geographical spatial information utilized for mapping and data analysis. As an example, the Water Resources District Office in Montana served real-time streamflow data to over 700 outside users via the World Wide Web (WWW) in September 1995.

The USGS has been very proactive in using electronic dissemination mechanisms. The Bureau has actively displayed a leadership role in Executive Order 12906, utilizing the National Spatial Data Infrastructure. The WWW is used extensively throughout the organization to provide access to USGS data. Publications formerly only available in print form are now available in an expanded form through the Web. The Survey also strives to meet the needs of its customers by making data available through a wide variety of media, including CD-ROM, 8-mm magnetic tape, 3480 cartridge tape, floppy disk, and CD-R.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
YES	YES	YES	YES	YES	YES	YES	YES
Free In-house	Free/Sale In-house	Free/Sale In-house	Free/Sale In-house	Free/Sale In-house	Free/Sale In-house	Free/Sale In-house	Free In-house
ESIC (800) 872-6277	FAX/AUTO (800) USA-MAPS	ESIC (800) 872-6277	Same	Same	Same	Same	9-track, 8 mm 3480 Cartridge Write-once (CD-R)

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

**U.S. Postal Service
(USPS)**

U.S. Postal Service
475 L' Enfant Plaza SW, Room 5821
Washington, DC 20260-2205

General Contact Information:

External Measurement Systems, 475 L'Enfant Plaza, SW, Room 5913, Washington, DC 20260-2205

The U.S. Postal Service provides quarterly estimates of first-class mail service performance into each of 96 cities from their respective aggregate overnight, two-day, and three-day service commitment areas, which are available in diskette form. Service performance data are currently collected by Price Waterhouse under contract to the U.S. Postal Service. Service performance is defined in two ways: the percentage of mail arriving within the service commitment and the average number of days to deliver. These data are in ASCII format. Dissemination of this information in diskette form began in 1993; no changes in format are anticipated.

Various national data about service performance for several classes of mail, customer satisfaction, and mail revenues and volumes are scattered, but available, on the Internet. Access USPS home page on the Internet: **<http://www.usps.gov>**, and select FYI, About the U.S. Postal Service, and then select either USPS Annual Report, Comprehensive Statement on Postal Operations, or Postal Facts.

Electronic Media Type Availability							
Electronic Bulletin Boards	Faxes & Auto Faxbacks	Direct Computer Dial-In	Internet	Optical Discs (CD-ROM)	Diskettes	Data Tapes	Other Electronic Media
NO	NO	NO	YES	NO	YES	NO	NO
			Free In-house www.usps.gov		Free In-house EMS		

Note: Free/Sale denotes whether a fee is charged; In-house/Outside describes where media are produced; bottom row contains contact information.

Reports Available in the Statistical Policy Working Paper Series

1. ***Report on Statistics for Allocation of Funds*** (Available through NTIS Document Sales, PB86-211521/AS)
2. ***Report on Statistical Disclosure and Disclosure-Avoidance Techniques*** (NTIS Document Sales, PB86-211539/AS)
3. ***An Error Profile: Employment as Measured by the Current Population Survey*** (NTIS Document Sales PB86-214269/AS)
4. ***Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics*** (NTIS Document Sales, PB86-211547/AS)
5. ***Report on Exact and Statistical Matching Techniques*** (NTIS Document Sales, PB86-215829/AS)
6. ***Report on Statistical Uses of Administrative Records*** (NTIS Document Sales, PB86-214285/AS)
7. ***An Interagency Review of Time-Series Revision Policies*** (NTIS Document Sales, PB86-232451/AS)
8. ***Statistical Interagency Agreements*** (NTIS Documents Sales, PB86-230570/AS)
9. ***Contracting for Surveys*** (NTIS Documents Sales, PB83-233148)
10. ***Approaches to Developing Questionnaires*** (NTIS Document Sales, PB84-105055/AS)
11. ***A Review of Industry Coding Systems*** (NTIS Document Sales, PB84-135276)
12. ***The Role of Telephone Data Collection in Federal Statistics*** (NTIS Document Sales, PB85-105971)
13. ***Federal Longitudinal Surveys*** (NTIS Documents Sales, PB86-139730)
14. ***Workshop on Statistical Uses of Microcomputers in Federal Agencies*** (NTIS Document Sales, PB87-166393)
15. ***Quality on Establishment Surveys*** (NTIS Document Sales, PB88-232921)
16. ***A Comparative Study of Reporting Units in Selected Employer Data Systems*** (NTIS Document Sales, PB90-205238)
17. ***Survey Coverage*** (NTIS Document Sales, PB90-205246)
18. ***Data Editing in Federal Statistical Agencies*** (NTIS Document Sales, PB90-205253)
19. ***Computer Assisted Survey Information Collection*** (NTIS Document Sales, PB90-205261)
20. ***Seminar on the Quality of Federal Data*** (NTIS Document Sales, PB91-142414)
21. ***Indirect Estimators in Federal Programs*** (NTIS Document Sales, PB93-209294)
22. ***Report on Statistical Disclosure Limitation Methodology*** (NTIS Document Sales, PB94-165305)
23. ***Seminar on New Directions in Statistical Methodology*** (NTIS Document Sales, PB95-182978)
24. ***Electronic Dissemination of Statistical Data*** (NTIS Document Sales, PB96-121629)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161; telephone: (703)487-4650. The Statistical Policy Working Paper series is also available electronically through the Bureau of Transportation Statistics World Wide Web home page (<http://www.bts.gov>).
